Boston College

The Graduate School of Arts and Sciences

Department of Physics

Symmetry and topology in condensed matter physics

a dissertation

by

XU YANG

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

MAY 2021

Symmetry and topology in condensed matter physics

XU YANG

Dissertation advisor: Dr. Ying Ran

Abstract

Recently there has been a surging interest in the topological phases of matter, including the symmetry-protected topological phases, symmetry-enriched topological phases, and topological semimetals. This thesis is aiming at finding new ways of searching and probing these topological phases of matter in order to deepen our understanding of them.

The body of the thesis consists of three parts. In the first part, we study the search of filling-enforced topological phases of matter in materials. It shows the existence of symmetry-protected topological phases enforced by special electron fillings or fractional spin per unit-cell. This is an extension of the famous Lieb-Schultz-Mattis theorem. The original LSM theorem states that the symmetric gapped ground state of the system must exhibit topological order when there's fractional spin or fractional electron filling per unit-cell. However, the LSM theorem can be circumvented when commensurate magnetic flux is present in the system, which enlarge the unit-cells to accommodate integer numbers of electrons. We utilize this point to prove that the ground state of the system must be a symmetry-protected topological phase when magnetic translation symmetry is satisfied, which we coin the name "generalized LSM theorem". The theorem is proved using two different methods. The first proof is to use the tensor network representation of the ground

state wave-function. The second proof consists of a physical argument based on the idea of entanglement pumping. As a byproduct of this theorem, a large class of decorated quantum dimer models are introduced, which satisfy the condition of the generalized LSM theorem and exhibit SPT phases as their ground states.

In part II, we switch to the nonlinear response study of Weyl semimetals. Weyl semimetals (WSM) have been discovered in time-reversal symmetric materials, featuring monopoles of Berry's curvature in momentum space. WSM have been distinguished between Type-I and II where the velocity tilting of the cone in the later ensures a finite area Fermi surface.To date it has not been clear whether the two types results in any qualitatively new phenomena. In this part we focus on the shift-current response ($\sigma_{shift}(\omega)$), a second order optical effect generating photocurrents. We find that up to an order unity constant, $\sigma_{shift}(\omega) \sim \frac{e^3}{h^2}\frac{1}{\omega}$ in Type-II WSM, diverging in the low frequency $\omega \to 0$ limit. This is in stark contrast to the vanishing behavior ($\sigma_{shift}(\omega) \propto \omega$) in Type-I WSM. In addition, in both Type-I and Type-II WSM, a nonzero chemical potential μ relative to nodes leads to a large peak of shift-current response with a width $\sim |\mu|/\hbar$ and a height $\sim \frac{e^3}{h}\frac{1}{|\mu|}$, the latter diverging in the low doping limit. We show that the origin of these divergences is the singular Berry's connections and the Pauli-blocking mechanism. Similar results hold for the real part of the second harmonic generation, a closely related nonlinear optical response.

In part III, we propose a new kind of thermo-optical experiment: the nonreciprocal directional dichroism induced by a temperature gradient. The nonreciprocal directional dichroism effect, which measures the difference in the optical absorption coefficient between counterpropagating lights, occurs only in systems lacking inversion symmetry. The introduction of temperature-gradient in an inversion-symmetric system will also yield non-

reciprocal directional dichroism effect. This effect is then applied to quantum magnetism, where conventional experimental techniques have difficulty detecting magnetic mobile excitations such as magnons or spinons exclusively due to the interference of phonons and local magnetic impurities. A model calculation is presented to further demonstrate this phenomenon.

Acknowledgments

First off, I wish to express my deepest gratitude to my advisor, Prof. Ying Ran. Numerous topics have I worked on and many extensive and intensive discussions with him are of the most joyous moments during my graduate years. His guidance in always asking the right question, in never taking any unexamined opinion for granted, has largely shaped my view of physics. Besides, his numerous advice on my career path and personal life have helped me greatly. I am truly lucky to have him as my advisor.

I would also like to thank the rest of my thesis committee: Prof. Ziqiang Wang, Prof. Ken Burch and Prof. Fazel Tafti. I thank Prof. Wang for many assistance during the past years including course selection, research proposal exam, etc.. I thank Prof. Burch for collaborations and solid state physics knowledge I have learned from him during various discussions. I thank Prof. Tafti for discussions and solid state physics I have learned from his class and also for being a committee member in my research proposal exam.

I am deeply thankful to Prof. Masaki Oshikawa for his hospitality during my visit in Japan, and for the many deep physics insights he has imparted to me and for his help during my postdoc application. And I would also like to thank Prof. Ashvin Vishwanath for the collaboration and for his leadership in the field of condensed matter physics which deeply affects my own research. I am also thankful to Prof. Gang Chen for collaborations, science discussions, and his hospitality during my visit in Hong Kong. I am also grateful to the faculty members of the Physics Department, especially, Prof. Michael Graf, Prof. Ilija Zeljkovic, Prof. Xiao Chen, Prof. P. Bakshi. I am also grateful to staffs in the department,

epecially Jane, Nancy, Sile and Scott. Their kindness to me has made my stay at Boston College a very pleasant experience.

I also want to thank my colleagues and collaborators, both at Boston College and elsewhere. Especially I would like to thank Shenghan Jiang, Kun Jiang, Tong Yang, Xiaodong Hu, Joshuah Heath and Hanlong Fang. The uncountable numbers of discussions on physics, metaphysics and pseudophysics and delicious meals shared with them have shaped both my mind and my body. I would also like to thank He Zhao, Zheng Ren, Lidong Ma, Yiping Wang, Hong Li at Boston College; and Xinqiang Cai, Zhicheng Yang, Yahui Zhang, Yuzi He, Yuwen Hu, Yuan Yao and Chunxiao Liu at other institutes, and my friends outside of physics, Yuansheng Zhou, Ting Zhu, Teng Bian, Zejie Yu, Zhichuang Sun, Siyi Ye, Qi Guo, Xiang Gao, Shiping Wang, Yongtao Wang, the happy moments spent with whom prevent me from getting permanent head demage (a.k.a, PHD).

Finally, I would like to thank my parents, whose enduring love and never-ceasing support make me become what I want to be. And I would like to thank God our Lord, "for you, O Lord, have made me glad by your work; at the works of your hands I sing for joy."[1]

[1]Psalm 92:4, ESV

To My Parents.

Contents

List of Figures

Chapter 1

General prologue

The best of artists hath no thought to show
which the rough stone in its superfluous shell
doth not include; to break the marble spell
is all the hand that serves the brain can do.

-Michelangelo

1.1 Overview of condensed matter physics

Condensed matter physics is a branch of physics that is dealing with condensed phases of matter. [1]

The first question one can ask in the field of condensed matter physics is: why are there so many different phases of matter? In the case of H2O, we know that it can be in the form of ice, liquid water and water vapor. And through the change of pressure and temperature, these different phases can transform into each other, with drastically different apperances. This curious fact even leads the Greek philosopher Thales to make the bold claim that

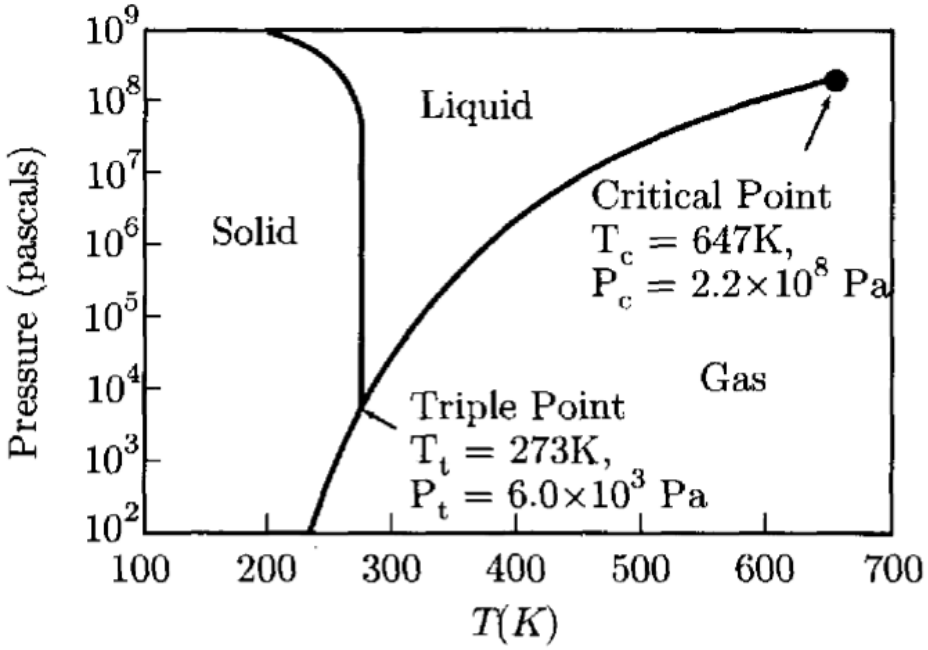


Figure 1.1: The $p - T$ phase diagram of water. The first-order liquid-gas transition line ends at the critical point with $T_c = 647K, P_c = 2.2 * 10^8 Pa$. The phase transition at the critical point becomes a second-order one, with continuous change of density and any other first order derivatives of the thermodynamic potential. Beyond the critical point, there is no phase transition between liquid water and water vapor.

everything is made of water. [2] Here's a twist: the difference between liquid water and water vapor is actually pretty vague. Usually one would differentiate between liquid water and water vapor through the process of evaporation, which is a first-order transition at which their densities have an abrupt jump. But the $p - T$ phase diagram shows that this transition line ends at one point, where the density-difference is zero, and beyond which there's no clear distinction between liquid water and water vapor. Are they truly different phases or just the same kind of phase? [1]

Another interesting point associated with phases of matter is as follows. We know from

our ordinary experience that liquid and gas are isotropic and uniform (in mathematical language it means that they are symmetric under the $SO(3)$ spatial rotation and continuous translation along three directions). At low enough temperature, usually the liquid will form a crystal, which breaks the $SO(3)$ rotation and continuous translation down to discrete rotations and discrete translation. This is indeed an astonishing effect, since we know that the law of electromagnetic force (the dominating force between the atomic scale and the everyday scale) is apparently isotropic. From quantum mechanics, we also know that the eigenstates of any Hamiltonian with symmetry G can always be made to be symmetric with respect to G [3]. There seems no reason for nature to choose a set of non-symmetric states as the basis in the degenerate space. Let's take the example of Ising symmetry-breaking in the transverse-field Ising model as an illustration. [5]

$$H = \sum_i -JS_i^z S_{i+1}^z - h_x S_i^x. \tag{1.1}$$

This model has the spin-z-flip symmetry $\prod_i S^x$. And when $h_x = 0$, the ground states are $|\uparrow\uparrow\cdots\rangle$ and $|\downarrow\downarrow\cdots\rangle$, which spontaneously breaks the spin-z-flip symmetry. This picture is not altered significantly when $h_x \ll J$ (below we call them $|\uparrow\rangle$ and $|\downarrow\rangle$ states). The skeptical might immediately object that the linear combinations $(|\uparrow\rangle \pm |\downarrow\rangle)/\sqrt{2}$ work just as well. In fact, for a finite system, the symmetric state has a lower energy than the anti-symmetric state and the symmetry-breaking phenomenon simply does not occur. The solution to this puzzle is that in the thermodynamic limit $N \to \infty$, the energy splitting between the symmetric state and the anti-symmetric state is of order h_x^{-N}, an exponentially small factor. Therefore they can be treated as degenerate safely. Similar consideration shows that any matrix elements of a local operator O between $|\uparrow\rangle$ and $|\downarrow\rangle$ states are zero in the thermodynamic limit. Therefore any local observations of a symmetric observable

can be described using either $|\uparrow\rangle$ or $|\downarrow\rangle$ states, together with a formal average over them. Furthermore, we note that the degeneracy of the states can be lifted by an infinitesmal external magnetic field h_z. As a result, we might treat the symmetry-breaking states as the true physical states in every real sense. [4]

L. Landau has developed the idea of symmetry and symmetry-breaking into a very general and powerful theory, which explains not only why there are symmetry breaking and symmetric phases, but also how phase transition happens between the two. [6] Landau's theory of second-order phase transition goes as follows. First, he assumes that near the symmetry breaking second-order phase transition, there's an order parameter that characterizes the symmetry breaking. In the example of the transverse-field Ising model, we can simply choose the average value of σ^z as the order parameter. The sign and the magnitude of its value denotes the direction and the degree of the symmetry-breaking. Landau's next observation is that the free energy is a functional of the order parameter field, usually expanded in low order polynomials of the order parameter field and its derivatives. The transformation rule of the order parameter under the symmetry group imposes restrictions on the form of the free energy. By minimizing the free energy functional over the order parameter field, we can obtain the value of the order parameter field in terms of the tunable parameters in the free energy functional, which are related to external conditions that can be tuned to trigger the phase transition. It can be clearly seen that in a continuous phase transition, the order parameter grows continuously from zero, signifying the phenomenon of symmetry-breaking.

Landau's theory of second-order phase transition shows that we can characterize phases of matter in terms of their symmetry properties. Therefore liquid water and water vapor are

essentially the same phase, but ice is a truly different phase since it has a lower symmetry than liquid water or water vapor. Landau's theory also shows continuous phase transition can only occur between two phases where the symmetry group of one phase is a subgroup of the symmetry group of the other phase. All the essential physics encoded in this phase transition can be described in terms of an order parameter field. Therefore from Landau we have a complete classification and understanding of physical states in terms of symmetry. The rest seems to systematically apply this machinary to all the known phases of matter. In fact, in the case of crystallography, there is the classification of crystals in terms of their different crystal symmetries, which is essentially working out all the point group symmetries compatible with a periodic array of atoms. We can then fit all the known crystals into this grand scheme. Without going into any detail, we already know that crystals with the same symmetry group share many physical properties in common. And the possible structural transition from one crystal into another crystal with higher or lower symmetry can be readily predicted using Landau's theory. [7] Yet this is not the whole story. As we shall see below, topology also plays an important role in the classification of phases of matter.

1.2 The advent of topological era

What we mean by topology is always associated with some kind of rigidity. The simplest example to demonstrate the phenomenon of topology is this famous joke: a topologist cannot tell the difference between a coffee mug and a donut, because they can be continuously deformed into each other without gluing or tearing. The rigidity lies in the fact that the number of holes is always the same during the deformation process since we do not al-

low gluing or tearing processes which are the only operations that can change the number of holes. [8] Solid state physics naturally provides us with such rigidity here and there, with various indications toward phenomena of topology. The rigidness of the Fermi surface topology is ensured by the Pauli exclusion principle-temperature only blur the Fermi surface by a very small degree at room temperature, therefore the whole Fermi surface topology is essentially unaltered. [10, 11] And the rigidness of the crystalline defects is ensured by the fact that an extensive amount of energy is needed to create or destroy a single crystalline defect. [9] In insulators, i.e., system with a energy gap to charge excitations, the rigidity is ensured by the relative difficulty of exciting a charged quasiparticle across the energy gap, and this is the case we are going to explore further in this section.

The modern era of topology in solid state physics begins with the following discoveries: the resonating-valence-bond state of quantum magnets, Berezinskii-Kosterlitz-Thouless transition, integer and fractional quantum Hall effects, the Haldane model and the spin-1 Haldane chain. [?, ?, ?, 12, 14–17] And the topological revolution reaches its climax with the discovery and systematic classification of quantum spin liquids, topological insulators and symmetry-protected topological phases in interacting bosonic systems. [18–23] These new discoveries show that there can be different phases even when the symmetries are exactly the same, and there can even be continuous phase transitions between them (e.g., BKT transition). Therefore an understanding of these phases certainly calls for a new perspective which encompasses the Landau paradigm.

Let's first take a closer look at the Landau paradigm to see what could possibly be missing. In the Landau paradigm, we have encoded all the relevant information of a state in terms of a uniform order parameter. For the symmetry-unbroken phase, we know that

the value of order parameter is zero. One can readily construct such a state as the direct product of identical wave-function which is a singlet under the symmetry group. In the case of transverse-field Ising model, we can model the symmetry-unbroken phase with $J = 0$ as the direct product of spins along the $+x$ direction, $|++\cdots\rangle$. Landau's theory tells us that all the other ground states under different values of J, h_z are basically "the same" as this simple direct product state, as long as no phase transition occurs. Here by "the same" we mean that the physical behavior are qualitatively the same, but can of course differ quantitatively (below we will try to put this hand-waving argument on a more solid ground). This line of reasoning can also be applied to the symmetry-breaking phase.

Therefore when applying Landau's theory of phase transition to the classification of phases of matter, one might draw a over-generalized conclusion that within every phase one can find a direct product state, which expresses the essential physical properties of the phase faithfully. But the new findings of topological phases show that this is definitely not the case. It is possible that there are some new states that has non-local information stored in the wave-function, which could not be described by a mere order parameter, and hence they behave drastically differently from a direct product state. Now it is a good time to explore further the idea of a phase. Below we shall restrict our discussion to quantum phase transition (mere convenience) and gapped phases of matter (gapless phases of matter are still not fully understood). States within the same gapped phase are "the same" in some sense, which can be made more precise by the idea of adiabatic evolution [24]. From the adiabatic theorem, we know that if the Hamiltonian depends on a parameter g and if g changes relatively slowly with time, then an eigenstate of $H(g)$ will stay as an eigenstate of $H(g)$ during the course of time evolution. The idea of adiabatic evolution then provides

us with the definition of a phase: if two gapped states $|\Psi_0\rangle$ and $|\Psi_1\rangle$ are in the same phase, then we can always find a family of Hamiltonian $H(g)$ with the tunable parameter $g \in [g_i, g_f]$, such that the energy gap for $H(g)$ are finite for all g, and the ground states of $H(g_i)$ and $H(g_f)$ are $|\Psi_0\rangle$ and $|\Psi_1\rangle$, respectively. This adiabatic time evolution is also equivalently called lcoal unitary evolution. From this new perspective, what we have said above can be reiterated as follows: all states in the same phase as a direct product state can be reached by proper local unitary evolutions, during which the gap of the Hamiltonians remain open, therefore the direct product state serves as a good representation of this phase. But the advent of the topological era tells us that even for systems with the same symmetry, we might have states that cannot be adiabatically connected to each other.

Let's first discuss the case where there's no symmetry present in the system. It turns out that there can be phases other than the conventional trivial phase. This is most clearly illustrated by the example of Kitaev's toric code model [25]. This model is an exactly-solvable spin model with not symmetry at all, and its 2d version has ground state degeneracy on high-genus Riemann surfaces (the simplest example being torus with genus-1, which naturally occurs if we impose periodic boundary conditions in the two spatial directions). This property is of particular interest since it directly reflects the topological structure of the real space configuration. The difference between the ground states of the toric code model and a direct product state is pretty clear, since the topological degeneracy between the 4 ground states of the toric code model on a torus can in no way be lifted by any local unitary transformation. Since ground state degeneracy usually results from some kind of symmetry breaking and the development of certain order, Xiao-gang Wen has drawn this analog and coined a name for such phases as "topological-ordered phases" [26]. This toric

code model also has other interesting features such as emergent Z_2 gauge field, emergent excitations with non-trivial mutual statistics and the emergence of fermionic excitations in a purely spin model, all of which are different incarnations of the underlying topological-ordered ground states. The role of quantum entanglement is also quite clear from the exact ground state wavefunction, which are a coherent superposition of macroscopic numbers of quantum states and can in no way be simplified by any adiabatic evolution of gapped Hamiltonians. This pattern of long-range entanglement of the ground state wave-function is in fact a characteristic feature of the topological ordered state.

Let's now discuss idea of adiabatic evolution in the presence of symmetry. Previously we have impose no restrictions on the Hamiltonian during the adiabatic time evolution other than the condition that gap is not closed. When the symmetry is present, however, it is necessary that at intermediate stages during the time evolution, the ground states are symmetric, so we need to require that the Hamiltonians during the evolution are symmetric. If we cannot find any symmetric adiabatic time evolution to connect two states with exactly the same symmetry and without topological order, we can say that these two states belong to two different phases of matter. Haldane phase and $S_z = 0$ phase of spin-1 chain are examples of states with the same symmetry which belong to two different equivalent classes of symmetric adiabatic time evolution. Band insulators and topological insulators are other examples. Note that symmetry is essential in the classification of these phases. If symmetry can be broken in intermediate steps, these states are in fact adiabatically connected to each other. Therefore they are termed "Symmetry-protected topological phases" (SPT). The above discussion also gives us a by-product: there are gapless modes on the boundary of a SPT phase, since if we view the vacuum as a trivial SPT state, then on the boundary

between these two different SPT states the gap must be closed for some modes.

Quantum spin liquids show an interesting interplay between symmetry and topology, specifically in the concept of symmetry fractionalization. When symmetry is present in the topological ordered states, we can discuss the symmetry properties of the topological excitations. Due to the fact that physical local operators never create or annihilate a single topological excitation, topological excitations always come in groups. In this sense we say that topological excitations are (in a sense) fractions of local excitations. In the same sense, the quantum numbers carried by topological excitations are also fractions of the symmetry quantum number of local excitations. This is best illustrated in the case of spinons in quantum spin liquids, which is a topological ordered state with $SO(3)$ spin rotation symmetry. Usually in a magnetic ordered state, there are magnons carrying spin-1 that can be created/annihilated by local spin flip operators. Heuristically, spinons in quantum spin liquids can be viewed as fractions of magnons, therefore they carry spin-1/2, which is a projective representation of the $SO(3)$ group. Symmetry fractionalization also occurs when other kinds of symmetry are present, such as time-reversal symmetry, crystal symmetries. These are topological-ordered states "enriched" by symmetry, since the topological order always exists no matter the presence or absence of symmetry. Therefore they are termed "Symmetry-enriched topological phases".

So far we have showed that the idea of classifying phases in terms of equivalence classes of local unitary evolutions w/o symmetry has included all the new phases beyond Landau paradigm, therefore providing us with a unified way of systematically classifying phases of matter.

Finally let me give a short remark on the experimental detections of the topological

phases. Since the topological nature of these phases are buried in their entanglement pattern of the wave-functions, the experimental detection of these novel phases of matter becomes a non-trivial task. The situation of symmetry-protected topological phases is slightly better, since general principle tells us that the boundary between such a material and the vacuum exhibit gapless modes [27]. There also exists other types of experiments, such as topological magnetoelectric effect in the case of topological insulators [28], etc.. One might ask if there are other experiments that can reveal the topological nature of the SPT phases. The situation of the symmetry-enriched topological phases is less promising, particularly because proper experimental probe is lacking. More is to be discussed on this point in the next section.

1.3 Structure of the thesis

Now I delineate the structure of my thesis. Chapter 2 is concerned with a generalized Lieb-Schultz-Mattis theorem. This is an attempt to set up a general guidance in the experimental search of SPT phases. The Lieb-Schultz-Mattis theorem, and its extension by Hastings and Oshikawa [30–32], can be stated as follows: if we have a system with fractional charge or fractional spin per unit-cell, the ground state of the system cannot be a symmetric gapped state without topological order. The ground state can be either one of the three alternatives: 1. it is a gapless state, 2. it breaks some symmetry, 3. it is a symmetry-enriched topological state (this is only possible in dimension > 1). The HOLSM theorem is a very useful guide in the field of quantum spin liquid. In a Mott insulator, we are given spin-1/2 per unit-cell. Suppose in experiments we do not detect any kinds

of symmetry breaking (spin-rotation, crystal symmetry, etc.), we can say that the ground state is most likely to be a quantum spin liquid.

On the face value, the HOLSM states the absence of a trivial state without symmetry breaking and without emergent gauge field. But given the data stated in the set-up, we can say more about the possible long-range ordered states. For example, in the case of square lattice with spin-1/2 per site, we can say that if the ground state is a gapped long-range ordered state with emergent gauge field, one of the gauge excitation must carry spin-1/2, i.e., it is a fractional excitation. The heuristic picture is as follows. The Mott insulator has a fractional spin per unit-cell. In order to keep the full translation symmetry and spin rotation symmetry in the ground state, we need to have spin-1/2 excitations per site to screen the background spin in the unit-cell. But no local excitation carries $S = 1/2$ (the most natural spin-flip excitations have spin-1), which means such excitaions must be topological excitations. [29]

From this new perspective, we find that HOLSM actually provides us with restrictions on the possible topological ordered states realizable in the system. Is there a similar theorem restricting possible short-range entangled state realizable in the system? This is the question posed and solved in Chapter 2. The solution is as follows. Starting from the HOLSM set-up, we know that there must be topological excitations carrying fraction spin to exactly screen the fractional spin per unit-cell in order to get a symmetric gapped ground state. But assume we further insert symmetric flux of symmetry g in each unit-cell (in the case of $U(1)$ charge symmetry, this is just a magnetic flux), we can have an alternative solution to the HOLSM constraint: the symmetry flux can provide us with the necessary fractional spin, thereby avoiding the occurrence of topological excitations. Such a state

must then be a non-trivial symmetry-protected topological phase, since in a trivial state (one that is adiabatically connected to vacuum), the symmetry flux of one group g does not possibly carry the fractional spin of another symmetry group ($SO(3)$ in this case). Under this general guidance, we consider 2+1D lattice models of interacting bosons or spins, with both magnetic flux andfractional spin in the unit cell. We propose and prove a modified Lieb-Shultz Mattis (LSM) theoremin this setting, which applies even when the spin in the enlarged magnetic unit cell is integral. The nontrivial outcome for gapped ground states that preserve all symmetries is that one necessarily obtains a symmetry protected topological (SPT) phase with protected edge states. This allows us to readily construct models of SPT states by decorating dimer models of Mott insulators to yield SPT phases, which should be useful in their physical realization. The resulting SPTs display a dyonic character in thatthey associate charge with symmetry flux, allowing the flux in the unit cell to screen the projective representation on the sites. We provide an explicit formula that encapsulates this physics, which identifies a specific set of allowed SPT phases.

Chapter 3 concerns the nonlinar photogalvanic response study of Weyl semimetals [33]. Recently, Weyl semimetals have been discovered in many materials with strong spin-orbit coupling. The topology of the electronic band structures gives rise to linear band touching points-Weyl nodes in momentum space, which are monopoles of the Berry's connection. These topological semimetals have been shown to host various exotic properties such as surface Fermi arcs, semi-quantized anomalous Hall effect, angle-dependent negative magnetoresistance, novel nonlinear optical effects. [34] The non-linear optical response has received increasing attention as a means to probe the Berry curvature of materials in general. This suggests non-linear optical effects can be used to distinguish between materials

with different Fermi surface topologies, a question particularly relevant to WSM. Indeed, shortly after the discovery of the first Type-I WSM material in TaAs, it was realized the tilt of velocity of the cone can be severe as to result in finite Fermi surfaces at all doping levels. [35] Nonetheless a clear distinguishing experimental consequence between these Type-II and their Type-I counterparts has yet to emerge.

The bulk photovoltaic effect (also called shift-current) is long studied in the field of semiconductors. It is an intrinsic second-order optic effect which converts light into electric currents. The microscopic mechanism of the BPVE can be heuristically understood as the change in polarization due to optical absorption, which can be readily represented in terms of covariant derivatives of Berry connections. Therefore this works as a direct probe of the Berry connections in the momentum space. This makes Weyl semimetal a natural platform for such a measurement due to the fact that Berry connection is divergingly large near the Weyl node.

The dimensional analysis shows that the BPVE response tensor $\sigma^{II}(\omega)$ should be $\frac{e^3}{h}$ times one over some energy scale. Naturally one would expect this energy scale to be just the energy of injecting photon. But a detailed calculation shows that this is only the case for type-II Weyl semimetal, i.e., $\sigma^{II}(\omega) \sim \frac{e^3}{h^2\omega}$. For type-I Weyl semimetal, however, we find that the leading contribution to the BPVE response is in fact proportional to ω. This we see as the fundamental difference between type-I and type-II Weyl semimetals. And the enhancement of BPVE signal in the $\omega \to 0$ limit in the type-II Weyl semimetal can be used as a detection of THz lights. Therefore the study of BPVE in type-II Weyl semimetal is of both theoretical and practical significance. In addition, in both Type-I and Type-II WSM, a nonzero chemical potential μ relative to nodes introduce a new energy scale, and

can be shown to lead to a large peak of shift-current response with a width $\sim |\mu|/\hbar$ and a height $\sim \frac{e^3}{h}\frac{1}{|\mu|}$, the latter diverging in the low doping limit. We show that the origin of these divergences is the singular Berry's connections and the Pauli-blocking mechanism. The second harmonic generation is also studied for the type I and type II Weyl semimetals, whose real part behaves similarly.

Chapter 4 studies the nonreciprocal directional dichroism in the field of quantum magnetism. The last chapter has shown the power of nonlinear electric responses in the field of topological semimetals. In this chapter, the idea is further explored by the study of nonlinear thermo-electomagnetic effect. The main motivation of this work is the call for proper experimental probes in the field of quantum magnetism. Novel states of matter in quantum magnets like quantum spin liquids attract considerable interest recently. Despite the existence of a plenty of candidate materials, there is no confirmed quantum spin liquid, largely due to the lack of proper experimental probes.

The existing experimental probes in this field can be roughly divided into three main categories:

1. Thermodynamics, including specific heat, magnetic susceptibility, etc.

2. Spectroscopy experiments, including neutron scattering, nuclear magnetic resonance, optical absorption, Raman scattering, etc.

3. Transport experiments, including electric conductivity, thermal conductivity, etc.

Ideally we would like to directly probe the mobile magnetic excitations in quantum magnets, such as magnons or spinons. Yet the traditional experiments do not probe the mobile magnetic excitations exclusively. For instance, spectrosocopy experiments like neutron

scattering receive contributions from disorder-induced local modes, while thermal transport experiments receive contributions from phonons. Here we propose a thermo-optic experiment which directly probes the mobile magnetic excitations in spatial-inversion symmetricand/or time-reversal symmetric Mott insulators: the temperature-gradient-induced nonreciprocal directional dichroism (TNDD) spectroscopy. This effect is defined as the difference in the optical absorption coefficient of the material between counterpropogating lights in the presence of a temperature gradient. Unlike traditional probes, TNDD directly detects mobile magnetic excitations and decouples from phonons and local magnetic modes. The microscopic formulation is established and the size of the effect is estimated using only basic quantities such as mean-free-path, gradient of temperature, strength of the spin-orbital coupling etc.. The contributions of non-magnetic modes and localized magnetic modes are estimated and can be shown to be safely ignored. A concrete microscopic calculation on Kagome lattice is performed to demonstrate this phenomenon.

Bibliography

[1] P. M. Chaikin, T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge 2000).

[2] Aristotle, *Metaphysics*(Hackett Publishing Company, Inc.; UK ed. edition, 2016).

[3] L. D. Landau, E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Butterworth-Heinemann, Oxford 1981).

[4] P. W. Anderson, *Basic Notions of Condensed Matter Physics* (Westview Press/ Addison-Wesley, 1997).

[5] S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge 2011).

[6] L. D. Landau, E. M. Lifshitz, *Statistical Physics* (Butterworth-Heinemann, Oxford 1980).

[7] C. Bradley, A. Cracknell, *The Mathematical Theory of Symmetry in Solids: Representation Theory for Point Groups and Space Groups* (Oxford University Press, 2010).

[8] A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, 2001)

[9] N. D. Mermin, Rev. Mod. Phys. **51**, 591 (1979).

[10] L. Van Hove, Physical Review **89.6**: 1189 (1953).

[11] A. M. Kosevich, Low Temperature Physics **30.2**, 97 (2004).

[12] P. W. Anderson, Materials Research Bulletin **8(2)**, 153 (1973).

[13] V. L. Berezinskii, Sov. Phys. JETP, **32(3)**, 493 (1971).

[14] J. Kosterlitz, D. Thouless, Journal of Physics C: Solid State Physics, **6(7)**, 1181 (1973).

[15] M. Cage, et al. *The Quantum Hall Effect* (Springer Science & Bussiness Media, 2012).

[16] D. Haldane, Physical review letters **61(18)** (2015).

[17] D. Haldane, Physics Letters A **93(9)**, 464 (1983).

[18] X. G. Wen, Physical Review B, **65(16)**, 165113 (2002).

[19] M. Hasan, C. Kane, Reviews of Modern Physics, **82(4)**, 3045 (2010).

[20] X. L. Qi, S. C. Zhang, Reviews of Modern Physics, **83(4)**,1057 (2011).

[21] X. Chen et al. Physical Review B, **87(15)**,155114 (2013).

[22] A. Kitaev, AIP conference proceedings, **1134(1)**,22 2009.

[23] A. Schnyder, Physical Review B, **78(19)**, 195125 (2008).

[24] X. Chen et al. Physical Review B, **82(15)**, 155138 (2010).

[25] A. Kitaev, Annals of Physics, **303(1)**, 2 (2003).

[26] X.G. Wen, Q. Niu, Physical Review B, **41(13)**, 9377 (1990).

[27] Y. Xia et al., Nature Physics, **5(6)**, 398 (2009).

[28] V. Dziom et al. Nature Communications, **8(1)**, 1 (2017).

[29] M. Zaletel, A. Vishwanath, Physical Review Letters, **114(7)**, 077201 (2015).

[30] E. Lieb et al., Annals of Physics, **16(3)**, 407 (1961).

[31] M. Hastings, Physical Review B, **69(10)**, 104431 (2004).

[32] M. Oshikawa, Physical Review Letters, **84(7)**,1535 (2000).

[33] X. Wan et al., Physical Review B, **83(20)**, 205101 (2011).

[34] N. Armitage et al., Reviews of Modern Physics, **90(1)**, 015001 (2018).

[35] A. Soluyanov, Nature **527(7579)**, 495 (2015).

Chapter 2

Dyonic Lieb-Schultz-Mattis theorem

2.1 Overview

The Lieb Shultz Mattis (LSM) theorem [1], appropriately generalized to higher dimensions [2–5], requires that a gapped spin system with fractional spin (eg. S=1/2) per unit cell possess excitations with fractional statistics (anyon) and fractional quantum numbers (topological order), if all symmetries (including lattice translations) are preserved. This has served as a powerful principle to diagnose exotic phases such as the fractional quantum Hall effect, and quantum spin liquids. Furthermore, in some cases the nature of the resulting topological order can be further constrained by the microscopic data [6, 7].

In recent years there has been an explosion of activity on symmetry protected topological (SPT) phases, which feature protected boundary modes although the bulk is short range entangled (SRE) and in contrast to the situation above, is free of anyon excitations. These include phases like topological insulators, which can be captured by free fermion models [8, 9], as well as intrinsically interacting phases [10–12] A natural question to ask is - are

there setting where the microscopic data alone would enforce an SPT phase, in a fashion analogous to the LSM theorem? If so, for a particular set of microscopic data, can we further characterize precisely which kinds of SPT orders are mandated?

These questions are answered in the present work. We show that SPT order *must* arise when the following conditions are met. The first ingredient is magnetic translation symmetry, that is an enlargement of the unit cell due to the non-commutativity of the primitive translation operations. Second, we require that the primitive unit cell (ignoring the non-commutativity) does not admit a trivial insulating phase. This is arranged by requiring a projective representation at each lattice site. Finally, we need some compatibility conditions between these two ingredients that allow, among other conditions, that the enlarged unit cell to be effectively at integer filling, what admits a short range entangled ground state. The latter is then shown to be an SPT. Furthermore for 2+1D bosonic systems we explicitly calculate the allowed SPTs compatible with the microscopic specifications. In addition we construct exactly soluble lattice models of this phenomenon to demonstrate the validity of our conclusions. This general principle should aid in the search for SPTs in realistic settings and exposes anew aspect of the interplay between symmetry and topology.

To give some simple plausibility arguments as to how microscopic details can enforce SPT order, consider free fermions in a magnetic field, when the filling fraction (ratio of particle density to magnetic flux density) is an integer. Then, an integer number of Landau levels will be filled, leading to a Chern insulator - which is a SRE topological phase with gapless edge states. Even in the presence of a lattice , one can establish a similar connection between the Hall conductance σ_{xy}, the flux n_ϕ and electron filling in the unit cell n_e [13, 14] which has been extended to the case of time reversal symmetric topological insulators [15].

To state our result more precisely, we consider a two dimensional lattice where the unit translations obey: $T_x T_y T_x^{-1} T_y^{-1} = g$, where g is an element of the symmetry group G. This generalizes the notion of a magnetic translation, particles acquire a phase factor depending on their g charge. We assume g is in the center of the symmetry group G (i.e. commutes with all other elements), but otherwise consider a general G, which can either be discrete or continuous, Abelian or nonAbelian, and can include time reversal implemented by an antiunitary representatation. Furthermore, in each unit cell a projective representation of the symmetry group labeled by 'α' is present. We derive a formula which provides a necessary and sufficient condition on these inputs to allow for a SRE phase, and determine constraints on the resulting SPT. Physically, this formula demands that a symmetry flux g inserted into this system will precisely generate a projective representation that can screen 'α' [15].

Let us give two physical pictures to view this filling and flux enforced SPTs. First we describe a vortex condensation based picture, for a system of lattice bosons with a conserved U(1) charge, with flux n_ϕ and filling n_b per lattice unit cell. Although our chapter focuses on having projective representations per site (rather than fractional filling) this example will be useful to build intuition. It is well known that a conventional insulator can be thought of as a condensate of vortices. However, for fractional filling n_b, the vortices see a fractional flux per unit cell [16], and their condensate will break lattice symmetries. Similarly, the bosons themselves cannot condense without breaking lattice symmetries due to the fractional flux n_ϕ. *However* the bound state of a vortex and p bosons may be able to propagate freely if: $n_b \pm p n_\phi \in \mathcal{Z}$ is an integer. The resulting object is a boson for p even which can then condense giving rise to a SRE and symmetric insulator. These are nothing

but the Bosonic Integer quantum Hall insulators at $\nu = n_b/n_\phi = p$ [12, 17, 18]. Note, here the condensing particle carries unit vorticity and hence the resulting insulator is free of topological order [19] and also preserves the U(1) symmetry since the condensing charge is attached to vorticity. A generalization of this result to include arbitrary symmetry groups is the main result of this chapter. An interesting exception occurs for $p = 1$, which is realized for example when one has bosons at half filling (or a projective representation of $U(1) \rtimes Z_2$), and a π flux in each unit cell. The doubled unit cell is at integer filling. At first sight it appears we can obtain an insulator by condensing the vortex-charge composite which sees no net flux in the unit cell. However, this composite is a fermion and cannot be condensed. This is also seen by a flux threading argument [14] that constrains such SRE phases to have $\sigma_{xy} =$ odd integer, which is impossible for a SRE topological phase of bosons [17,18]. Interestingly, this result continues to hold if the $U(1)$ is broken to a discrete symmetry as shown below.

A second perspective is to begin in a topologically ordered phase with fractionalized excitations and consider confining all exotic excitations by an appropriate anyon condensate. For example, for bosons at half filling, one could obtain toric code (Z_2) topological order where the e particle carries half charge [20]. The m particle however sees the fractional charge density as background flux and cannot condense while preserving spatial symmetries. This is the situation in the absence of magnetic translations, where the LSM theorem enforces topological order for gapped symmetric states. However, once we allow for magnetic translations with g charge, a way out to an SRE phase may become available. The m particle, bound to a g charge that sees the magnetic flux, forms a composite object that may condense uniformly and confine the topological order. At the same time, this leads

to an SPT phase since the condensing anyon carries nontrivial symmetry charge [21, 22]. Indeed this picture will allows us to construct models of such LSM enforced SPT phases as we describe below.

Before discussing construction of models, it may be helpful to give a few examples. Consider a system of degenerate doublets ("S=1/2") on sites of a square lattice. This site degeneracy may arise from spin rotation invariance ($SO(3)$), or even just as Kramers degeneracy protected by time reversal Z_2^T symmetry. Now consider an additional Z_2 symmetry which is invoked in defining the magnetic translations, i.e. we have a fully frustrated Ising model on the same lattice. According to our results, in both these situations SRE ground states are possible but must be SPT phases. While the SPT phase is unique for the second case of Kramers doublets of $Z_T \times Z_2$, in the former case of $SO(3) \times Z_2$ there is more than one SPT phase possible. Interestingly, if we consider a minor modification of the $Z_T \times Z_2$ model, such that the doublets on each site are non Kramers pairs, protected by the combination of the two symmetries, then *no* SRE ground state exists (and hence no SPT exists) that respects all symmetries. These examples are discussed in detail in Section 2.3.1 which also introduces models that realize them.

In constructing models, the first step is to begin in the deconfined phase of a discrete lattice gauge theory (or of a dimer model). Then, one way to obtain a confined phase is by decorating the electric field lines with domain walls of a global symmetry. This identification implies that we have condensed the composite of magnetic flux and symmetry charge. The resulting confined phase is potentially an SPT if the electric charges are associated with the appropriate symmetry fractionalization [21, 22]. However, to obtain an LSM enforced SPTs the situation is different since they involve fractional spin on the sites. In a dimer

model this corresponds to having an odd number of dimers associated with a unit cell, in which case we cannot decorate them with regular domain walls (which should be closed loops). *However* if the global symmetry is also associated with flux in the unit cell (for example a fully frustrated Ising model), the two kinds of frustration cancel each other out, and one can still achieve this decoration of electric field line. This is discussed explicitly in Section 2.2, for a specific model and the resulting state is shown to be the desired SPT. The model there is one of hardcore bosons on the Kagome lattice tuned to half filling by particle hole symmetry, previously introduced by Balents Fisher and Girvin [23]. While their focus was on a Z_2 spin liquid phase, we decorate their model with an additional Z_2 symmetry realized by a fully-frustrated Ising model. The combination is shown to realize an LSM enforced SPT phase with gapless edge states, but a short range entangled bulk.

Finally in Section 2.3 we discuss the problem for general symmetry groups, and derive the necessary and sufficient conditions for SRE phases to emerge and identify the class of SPTs that must be realized. Proofs can be found in the appendices.

2.2 A simple model realizing SPT phase

Our discussion starts from a concrete microscopic model realizing an SPT phase. The beauty of this model is its simplicity, which only includes two-spin and three spin interactions. It turns out that the crucial features of this model can be systematically generalized which form the main results of the current study.

The model constructed below (see Eq.(2.5)) is based on the Balents-Fisher-Girvin(BFG) model [23]. The original BFG model [23] is a model with spin-1/2 residing on Kagome

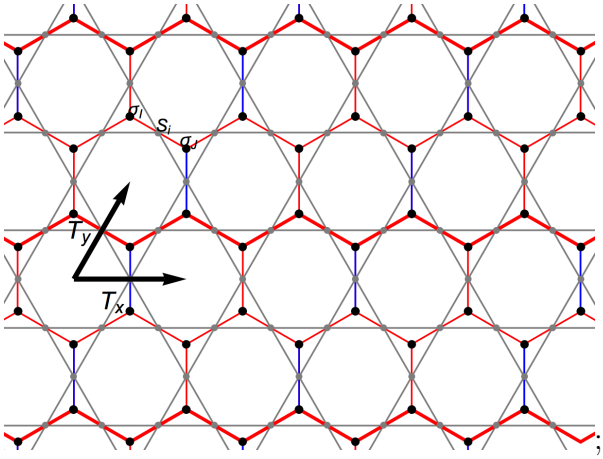


Figure 2.1: (color online) Degrees of freedom in the decorated BFG model. The Ising d.o.f. σ_I live on the honeycomb lattice and the spin d.o.f. S_i lives on the Kagome lattice. The Ising coupling signs $s_{IJ} = +1$ on red bonds, and $s_{IJ} = -1$ on blue bonds. The thick red bonds represent the "y-odd zigzag chains" used in Eq. (2.14)

lattice. It is the low energy effective Hamiltonian if we take the $J_z \gg J_\perp$ limit of the following XXZ Hamiltonian

$$H^{XXZ} = J_\perp \sum_{\hexagon} [(\sum_{i\in\hexagon} S_i^x)^2 + (\sum_{i\in\hexagon} S_i^y)^2 - 3] + J_z \sum_{\hexagon} (\sum_{i\in\hexagon} S_i^z)^2, \tag{2.1}$$

which has a spin-liquid ground state for $J_z \gg J_\perp$ with deconfined spinons as confirmed by various numerical methods [24, 25].

Let's then take a look at the low energy effective Hamiltonian. The limit $J_z \gg J_\perp$ ensures that $S^z_{\hexagon} = 0$ for every hexagon and the resulting Hamiltonian in this low energy manifold takes the following ring-exchange form

$$H^{BFG} = -J_{\text{ring}} \sum_{\bowtie} (| \quad \rangle\langle \quad | + h.c.), \tag{2.2}$$

with $J_{ring} = J_\perp^2/J_z$.

Let's then decorate the XXZ model by putting a layer of Ising spins σ inside every

triangle of the Kagome lattice, which comprises a honeycomb lattice. The Ising spins are in a transverse field, *i.e.*,

$$H^{Ising} = h \sum_I \sigma_I^x, \tag{2.3}$$

with Ising spin σ_I living on the honeycomb lattices labeled by I.

We then couple these two layers through a binding term

$$H^{binding} = - \sum_{\text{I } \overset{\text{i}}{\bullet\!\!-\!\!\bullet} \text{ J}} \lambda S_i^z \cdot (s_{IJ} \sigma_I^z \sigma_J^z), \tag{2.4}$$

where the summation is over all the bonds IJ on honeycomb lattice with S_i at the bond center. The sign $s_{IJ} = \pm 1$ are frustrated in the sense that $\prod_{I,J \in \hexagon} s_{IJ} = -1$. We have specifically chosen a choice of s_{IJ} in Fig. 2.1. The binding term binds spin-up with Ising happy bond ($s_{IJ}\sigma_I^z\sigma_J^z = +1$) and spin-down with Ising un-happy bond ($s_{IJ}\sigma_I^z\sigma_J^z = -1$).

The full Hamiltonian we are considering is then given by (see Fig. 2.1)

$$H = H^{XXZ} + H^{binding} + H^{Ising}. \tag{2.5}$$

One can divide H into two parts

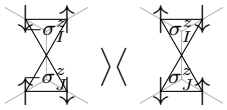
$$\begin{aligned} H_0 &= J_z \sum_{\hexagon} (S_{\hexagon}^z)^2 - \sum_{\text{I } \overset{\text{i}}{\bullet\!\!-\!\!\bullet} \text{ J}} \lambda S_i^z (s_{IJ} \sigma_I^z \sigma_J^z). \\ H_1 &= J_\perp \sum_{\hexagon} [(S_{\hexagon}^x)^2 + (S_{\hexagon}^y)^2 - 3] + \sum_I h \sigma_I^x. \end{aligned} \tag{2.6}$$

Considering the the limit where $J_z, \lambda \gg J_\perp, h$, we can first deal with H_0 and then treat H_1 as a perturbation. All the terms in H_0 commutes with each other and hence all the eigen-states and eigen-energies are known for H_0. In fact, there is a two-to-one mapping from the ground state sector to the low energy sector of the BFG model (i.e. $\{S_i^z\}$ configurations satisfying 3 $S^z = +1/2$ per hexagon). We will consider periodic boundary conditions, and

the Hilbert space of the original BFG model has four topological sectors labeled by parities of the $\prod_{k\in\mathcal{C}} 2S_k^z$ around the non-contractable loops $\mathcal{C}$ (which is just the non-contractable vison flux line [26]). This mapping only map onto one specific topological sector since $\prod_{k\in\mathcal{C}} 2S_k^z$ is identified with $\prod_{IJ\in\mathcal{C}} s_{IJ}$ due to $H^{binding}$. The preimage of any low energy $\{S_i^z\}$ configuration inside this topological sector are two states $|\{S_i^z,+\}\rangle$ and $|\{S_i^z,-\}\rangle$ (related to each other by a global Ising flip).

It turns out that the effective Hamiltonian in the parameter regime where $J_z \gg \lambda \gg J_\perp, h$ and $\frac{h^2}{\lambda^2} \gg \frac{J_\perp}{J_z}$ has the following form (see Appendix. 2.7.1 for detailed calculation)

$$H^{deco.BFG} = -\frac{10J_\perp^2 h^2}{9J_z\lambda^2}\sum_{\bowtie}(\Big| \quad \Big\rangle\Big\langle \quad \Big| + h.c.). \tag{2.7}$$



Note that the kinetic term in this effective Hamiltonian is the ring exchange term of four spins at the ends of each bowtie as in the original BFG model combined with the flipping term of the two Ising d.o.f. within this bowtie, such that the constraint $S_i^z(s_{IJ}\sigma_I^z\sigma_J^z) = 1$ is still satisfied everywhere.

We shall then prove that the ground state of the decorated BFG model is a symmetric short-range entangled SPT state. In fact, using the mapping $\mathcal{P}$ between the Hilbert space of the decorated BFG model and the original BFG model

$$\mathcal{P} : (|\{S_i^z,+\}\rangle + |\{S_i^z,-\}\rangle)/\sqrt{2} \to |\{S_i^z\}\rangle\,. \tag{2.8}$$

Such a mapping is clearly an isometry. Next we notice that

$$\mathcal{P}H^{deco.BFG}\mathcal{P}^{-1} = H^{BFG}, \tag{2.9}$$

with the identification $J_{ring} = \frac{10J_\perp^2 h^2}{9J_z\lambda^2}$, which can be proven by directly comparing the matrix elements on the two sides.

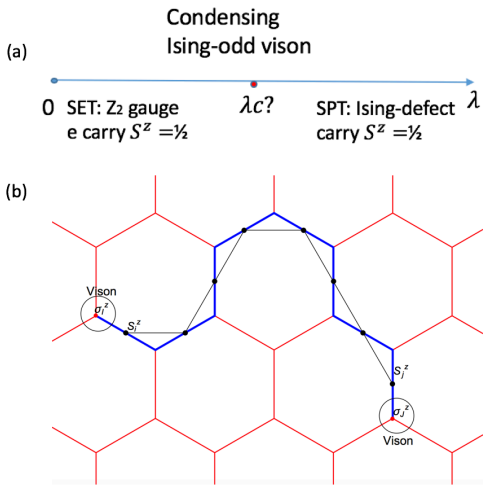


Figure 2.2: (color online) (a) Schematic phase diagram of the decorated BFG model by tuning λ. We have already fixed $J_z \gg J_\perp, h$. In the limit $\lambda \to 0$ the Ising layer is decoupled and the ground state is just that of the original BFG model with Z_2 topological order. This is an SET state with spinon carrying $S^z = 1/2$. When λ is tuned to be within the parameter regime where $J_z \gg \lambda \gg J_\perp, h$ and $\frac{h^2}{\lambda^2} \gg \frac{J_\perp}{J_z}$, we have an SPT state with Ising defect carrying $S^z = 1/2$ as discussed in the main text. There is a possible direct phase transition triggered by the condensation of Ising-odd visons at some intermediate λ_c. (b) A schematic view of vison condensation. The honeycomb lattice where Ising d.o.f. lives is shown and the spin d.o.f. lives in the bond center. Two visons are created at the ends I, J of the string operator $\sigma^z_I \sigma^z_J \prod_{k=i}^{j} 2S^z_k$. with S^z_k runs over all the black dot shown in the graph. Alternatively we can view the string operator as the product of bond variable $S^z_i \sigma^z_I \sigma^z_K$ along the thick blue bonds. Due to the constraint $S^z_i(s_{IK}\sigma^z_I\sigma^z_K) = 1$, the string operator will yield a factor (product of s_{IK}'s along the thick blue bonds) when acting on the ground state wave-function, which means the visons are condensed and the topological order is killed. Note that the condensed visons in the present case are dressed by local σ^z operator and hence carry the quantum number of Z_2 Ising symmetry, which result in an SPT state.

Therefore the spectrum of $H^{deco.BFG}$ within the Ising-even sector is exactly the same as that of H^{BFG} inside a specific topological sector, which is known to be gapped. And the ground state $|\psi\rangle$ of H^{BFG}, should also be mapped to the ground state $|\psi^{deco.}\rangle$ of $H^{deco.BFG}$. However there is still one possibility that there exists a state in the Ising-odd sector with exactly the same energy as $|\psi^{deco.}\rangle$, which features the Ising symmetry breaking. This possibility is ruled out because $|\psi^{deco.}\rangle$ has no long-range order in σ^z as will be discussed in the context of vison condensation.

The ground state of the original BFG model has Z_2 topological order which supports vison and spinon excitations. In the original BFG model, the vison excitations live in the honeycomb lattice and are created at the ends of the string of S_k^z operators [23], *i.e.*,

$$v_I v_J = \prod_{k=i}^{j} \rightarrow 2S_k^z, \tag{2.10}$$

where two visons are created at I and J (see Fig. 2.2).

We can see that the visons are condensed in $|\psi^{deco.}\rangle$. The vison operator $v^{deco.}$ at site i should now be dressed by the local Ising operators σ_I^z with I around i to obtain

$$v_I^{deco.} v_J^{deco.} = \sigma_I^z \sigma_J^z \prod_{k=i}^{j} \rightarrow 2S_k^z. \tag{2.11}$$

With the constraint that $S_i^z(s_{IJ}\sigma_I^z\sigma_J^z) \equiv 1$ and the fact that intermediate σ^z squared to 1, we know that $v_I^{deco.} v_J^{deco.}$ must yield a constant (depending only on the product of s_{IJ}'s along the vison string) when acting on $|\psi^{deco.}\rangle$, see Fig. 2.2 for an illustration.

Now it is clear that the correlator $\sigma_I^z\sigma_J^z$ is short-range because we have

$$\begin{aligned} |\langle\psi^{deco.}|\sigma_I^z\sigma_J^z|\psi^{deco.}\rangle| &= |\langle\psi^{deco.}|\prod_{k=i}^{j} \rightarrow 2S_k^z|\psi^{deco.}\rangle|, \\ &= |\langle\psi|\prod_{k=i}^{j} \rightarrow 2S_k^z|\psi\rangle|, \end{aligned} \tag{2.12}$$

where the last correlator exhibits exponential decay since visons are deconfined in the original BFG model (the last equality holds because $\mathcal{P}$ commutes with the string operator).

The above discussions feature the physical picture of the condensation of visons carrying Ising quantum number, which kills the Z_2 topological order. We will soon show that the resulting phase is an SPT phase, which is exactly a realization of the anyon condensation mechanism to obtain SPT phases proposed in Ref. [21]. In fact, if we start from the decoupling limit with $\lambda = 0$ and gradually increase λ with all other couplings fixed, we should be able to see two phases: an SET phase with spinon carrying $S^z = 1/2$ when λ small and an SPT phase resulting from the condensation of Ising-odd visons when λ is in the parameter regime $J_z \gg \lambda \gg J_\perp, h$ and $\frac{h^2}{\lambda^2} \gg \frac{J_\perp}{J_z}$. This two phases might be related by a continuous phase transition at some intermediate λ_c. See Fig. 2.2 for an illustration.

One way to see that the ground state of the decorated BFG model Eq.(2.7) is an SPT phase is to consider the Ising defects, which turn out to be topologically bound with $S^z = \pm 1/2$ — a projective representation of the symmetry group (see discussion below). In order to introduce Ising defects we need to take a branch cut and modify the terms straddling the branch cut such that only one side is conjugated by the Ising symmetry σ^x. The net effect is that for the bonds IJ crossed by the branch cut, the sign of s_{IJ} is flipped. See Fig. 2.3 for an illustration. To compute the S^z quantum number carried by the Ising defect it is convenient to introduce the equivalent hard-core boson description: $n_i \equiv S^z_i + 1/2$. Now let's take a loop $\mathcal{C}$ enclosing one of the Ising defect and measure the total charge within the area $\mathcal{D}$ bounded by $\mathcal{C}$. This is done by the following U(1) transformation

$$\prod_{\hexagon \in \mathcal{D}} \left(\prod_{i \in \hexagon} e^{i\frac{\theta}{2} n_i} \right) = \prod_{i \in \mathcal{C}} e^{i\frac{\theta}{2} n_i} \cdot \prod_{i \in \mathcal{D}/\mathcal{C}} e^{i\theta n_i}, \tag{2.13}$$

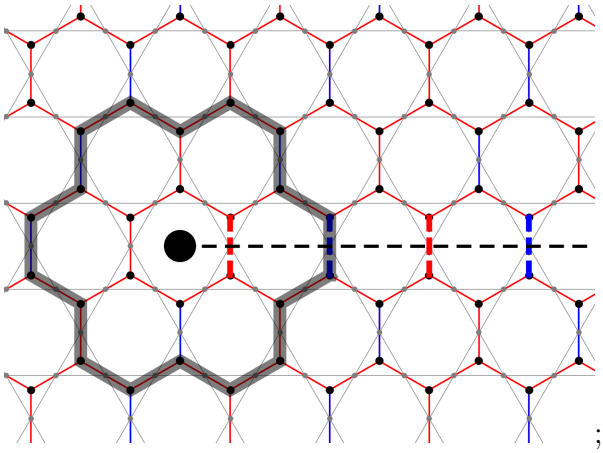


Figure 2.3: (color online) An pair of Ising defects (only one is shown) is created at the end points of the branch cut (dashed black line) after modifying the original Hamiltonian H in Eq.(2.5) into H'. The sign s_{IJ} is flipped in H' along the branch cut comparing with the original model. (red bond: $s_{IJ} = +1$, blue bond: $s_{IJ} = -1$) For any loop $\mathcal{C}$ enclosing the Ising defect as the gray loop shown here, the product $\prod s_{IJ}$ around the loop flips sign comparing with the original model. In order that $H'^{binding}$ does not cost extra energy, the spin should be flipped wherever s_{IJ} changes its sign. As a result, the total S^z around the loop $\mathcal{C}$ is changed by an odd integer. The result is that Ising defect is topologically bound with a half-integer spin. See the discussion in the main text.

from which we know that the fractional charge part is determined by boson numbers on the boundary only.

In order that the binding term after modification $H'^{binding}$ does not cost energy, the boson number n_i on the bond IJ across the branch cut should be changed by 1. Since $\mathcal{C}$ only crosses Ising defect line odd times, the total boson numbers around $\mathcal{C}$ should be changed by $1(mod\ 2)$. From Eq. (2.13), this amounts to the change of charge within $\mathcal{D}$ by $\frac{1}{2}(mod\ 1)$. This fact doesn't depend on the position of the branch cut or the position of $\mathcal{C}$ we have chosen, indicating that the extra $\frac{1}{2}$ charge is bounded with the Ising defect. In the original spin language it means Ising defect carries spin $S^z = \pm 1/2$.

Summary: Before proceeding let's pay close attention to the symmetry property of the Hamiltonian in Eq. (2.5). This model has the following $Z_{2g} \times (U(1) \rtimes Z_{2h})$ onsite symmetry

1. Spin-rotation symmetry $U(1)$: $\prod_i e^{i\theta(S_i^z + \frac{1}{2})}$, with $\theta \in [0, 2\pi)$.

2. Spin-flip and Ising-flip symmetry $Z_{2h} = \{I, h\}$: $\prod_i S_i^x \prod_{I \in A} \sigma_I^x$, which flips all the spins on Kagome lattice and Ising d.o.f on A sublattice of the honeycomb lattice. This symmetry operation leaves the Hamiltonian in Eq. (2.5) invariant.

3. Ising symmetry $Z_{2g} = \{I, g\}$: $\prod_I \sigma_I^x$, which flips all the Ising d.o.f. on the honeycomb lattice.

The onsite symmetry group has the direct product structure $G = Z_{2g} \times (U(1) \rtimes Z_{2h})$. In addition, every unit cell (three kagome sites and two honeycomb sites) carries a nontrivial projective representation of $U(1) \rtimes Z_{2h}$ (i.e. a half-integer spin). Naively one would suspect that the LSM theorem would rule out a symmetric short-range-entangled state in such a system. However, due to the frustrated nature of the gluing term, we actually have magnetic

translation symmetry T_x, T_y in the system instead of usual translation $T_x^{orig.}, T_y^{orig.}$, which can be written as

$$\begin{aligned} T_x &= (\prod_{I \in \text{y-odd zigzag chains}} \sigma_I^x) T_x^{orig.}, \\ T_y &= T_y^{orig.}, \end{aligned} \tag{2.14}$$

which satisfy the magnetic translation algebra

$$T_x T_y T_x^{-1} T_y^{-1} = g. \tag{2.15}$$

Therefore it is still possible for us to have an SPT state. And we also know that the g-defect carries $S^z = \pm 1/2$, which has the same projective representation as that carried by the unit-cell. Below we will find that this is not merely a coincidence and there is a deep connection between the patterns of short-range entanglement and the projective representation carried by a unit-cell related by the magnetic translation algebra.

2.3 Main Results

Our main results are captured in two theorems. Theorem-I is easier to state but less general. Theorem-II is more general but is more mathematically involved to state.

Consider a two-dimensional *bosonic* quantum system respecting an onsite symmetry group G (which could contain time-reversal), and g is a unitary symmetry element in the center of G (i.e., g commutes with any element in G). The system respects a "magnetic" translation symmetry group generated by T_x, T_y satisfying the algebra:

$$T_x T_y T_x^{-1} T_y^{-1} = g, \tag{2.16}$$

where T_x, T_y are assumed to be the usual translation operations combined with certain site-dependent onsite unitary transformations. We further assume that the physical degrees of

freedom (d.o.f.) in each real space unit cell (*not* the enlarged magnetic unit cell) form a nontrivial projective representation α of G, specified by a 2-cocycle: $\forall a, b \in G, \alpha(a, b) \in U(1)$ and $\alpha \in H^2(G, U(1))$. Precisely, the unitary or antiunitary transformation U_a, U_b of $a, b \in G$ satisfy:

$$U_a {}^a U_b = \alpha(a, b) U_{ab}, \tag{2.17}$$

where the left-superscript a in ${}^a U_b$ denotes the group action of a on U_b: if a is unitary (antiunitary), then ${}^a U_b = U_b$ (${}^a U_b = U_b^*$, i.e. complex conjugation of U_b).

We ask the following question: is it possible for such a system to have a short-range entangled(SRE) gapped ground state without breaking symmetries?

Here we use the definition of SRE states following Ref. [12]; i.e., those are gapped quantum phases that can be deformed into the trivial product state via local unitary transformations. Note that if the system respects usual translational symmetries, this would be impossible: constrained by a generalized Lieb Shultz Mattis theorem [1, 2, 4], the nontrivial projective representation per unit cell indicates that a gapped liquid ground state necessarily features topological order. Here because the system respects a magnetic translation symmetry, it is possible that a SRE liquid ground state exists. We give sufficient and necessary conditions for such a liquid phase to exist, and show that this SRE liquid phase must be an SPT phase.

In the presence of an onsite symmetry group G, focusing on bosonic systems specified by Eq.(2.16,2.17), we have the following theorems:

Theorem-I: Here we further assume that $G = G_1 \times Z_N$ where Z_N is the finite abelian subgroup generated by g. The quantum system above can have a SRE liquid ground state if and only if the two conditions below are both satisfied. Such a liquid phase is necessarily a nontrivial SPT phase because the g-symmetry defect must carry the projective representation α.

1. $\alpha^N \simeq \mathbf{1} \in H^2(G, U(1))$, i.e. N of these projective representations fuse into a regular representation.

2. The group function (which maps elements of the symmetry group to phases, while preserving the group relations) $\gamma_g^\alpha(a) \equiv \frac{\alpha(g,a)}{\alpha(a,g)}$, $\forall a \in G$ is a trivial 1-cocycle (or equivalently, a trivial one-dimensional representation):
   $\gamma_g^\alpha \simeq \mathbf{1} \in H^1(G, U(1))$.

What is the physical meaning of these two conditions? The first condition ensures that the enlarged magnetic unit cell does not have projective representations. Interestingly, this is not sufficient to ensure an SRE phase. Additionally, condition 2 must be satisfied, which essentially states that the symmetry involved in magnetic translations, g, can be chosen to commute with all other projective group actions in a proper gauge.

Theorem-II: Here we do not make extra assumptions on G. The quantum system above can have a SRE liquid ground state if and only if there exists a 3-cocyle ω_0: $\forall a, b, c \in G, \omega_0(a, b, c) \in U(1)$ and $\omega_0 \in H^3(G, U(1))$, such that the group function $\delta_g^{\omega_0}(a, b) \equiv \frac{\omega_0(g,a,b)\omega_0(a,b,g)}{\omega_0(a,g,b)}$ is 2-cycle equivalent to α^{-1}: $\delta_g^{\omega_0} \simeq \alpha^{-1} \in H^2(G, U(1))$. Such

a liquid phase is necessarily a nontrivial SPT phase because the g-symmetry defect must carry the projective representation α. The possible nontrivial SPT phases form a coset from the classification point of view (see Remark below).

Remark: it is straightforward to show that $\forall\alpha \in H^2(G,U(1))$, $\gamma_g^\alpha \in H^1(G,U(1))$. And similarly $\forall\omega \in H^3(G,U(1))$, $\delta_g^\omega \in H^2(G,U(1))$. The mappings $\gamma_g : H^2(G,U(1)) \to H^1(G,U(1))$ and $\delta_g : H^3(G,U(1)) \to H^2(G,U(1))$ reducing a n-cocycle to a $(n-1)$-cocycle are the so-called slant-products in mathematical context. γ_g and δ_g preserve the multiplication relation in the cohomology group. In particular, there is a subgroup $\mathcal{A}_g \in H^3(G,U(1))$ such that $\forall\omega \in \mathcal{A}_g, \delta_g^\omega \simeq \mathbf{1} \in H^2(G,U(1))$, (i.e., $\mathcal{A}_g$ the kernal of the mapping δ_g).

When the condition in Theorem-II is satisified, ω_0 must be a nontrivial element in $H^3(G,U(1))$ because α is nontrivial by assumption. And the realizable SPT phases form a coset from the classification point of view. More precisely, the 3-cocyle characterizing the SRE liquid phase must be one of the element in the following coset: $\omega_0 \cdot \mathcal{A}_g$.

Outline of the proof: The proof of these theorems is a combination of a pumping argument of entanglement spectra and derivations/constructions based on a recently developed symmetric tensor-network formulation [21], which we outline here. Basically, if a SRE liquid phase exists, by the pumping argument of entanglement spectra one knows that *the g-symmetry-defect in this phase must carry the projective representation α, and consequently this phase must be an SPT phase.* This physical observation can be further justified by calculations based on symmetric tensor-networks, leading to the following mathematical result: if the 3-cocycle characterizing the SRE liquid phase as $\omega \in H^3(G,U(1))$, then

magnetic translation symmetry dictates $\delta_g^\omega \simeq \alpha^{-1}$, which is exactly the same mathematical condition for the g-symmetry-defect carrying the projective representation α. In addition, based on the symmetric tensor-network formulation, for any ω satisfying $\delta_g^\omega \simeq \alpha^{-1}$, a SRE liquid phase characterized by ω respecting the magnetic translation symmetry can be constructed. These prove that the conditions in Theorem-II are *necessary and sufficient* for the SRE liquid phase to exist. In addition, when ω_0 exists, because a 3-cocycle $\omega \in H^3(G, U(1))$ satisfies $\delta_g^\omega \simeq \alpha^{-1}$ if and only if $\omega \in \omega_0 \cdot \mathcal{A}_g$, the coset structure in the Remark is also established.

Theorem-I is just a special case of Theorem-II. Namely when $G = G_1 \times Z_N$, one can show that if and only if the two conditions in Theorem-I is satisified does the condition in Theorem-II is satisfied. The condition-(1) in Theorem-I is well anticipated. If condition-(1) is not satisfied, then physical degrees of freedom form a nontrivial projective representation of G even in the enlarged magnetic unit cell (N times larger than original unit cell), and the generalized Lieb Shultz Mattis theorem [1,2,4] already forbids a SRE liquid phase to exist. The condition-(2) is less obvious and more interesting, which puts additional constraints for the existence of a SRE liquid phase (see example-(4) below).

Before going into the details of the proof, let us consider a few simple examples to see the applications of the Theorems and the Remark.

2.3.1 Examples

In these examples, the element g in the magnetic translation algebra Eq.(2.16) generates a $Z_2^{Ising} \equiv \{I, g\}$ Ising symmetry group. For instance, a fully frustrated Ising model on the square lattice would satisfy this magnetic translation symmetry. The symmetry-enforced

SPT phases in example-(1,2,3) will be demonstrated via a class of decorated quantum dimer models, which are exactly solvable at the Rokhsar-Kivelson points [27].

(1) $\mathbf{G = SO(3) \times Z_2^{Ising}}$, and a spin-1/2 per unit cell: Namely, the projective representation α per unit cell is nontrivial because only the $SO(3)$ part is projectively represented, and the Ising and the spin-rotation still commute: $\alpha(g, a) = \alpha(a, g),\ \forall a \in G$. Clearly the two conditions in Theorem-I are both satisified. First, two spin-1/2's fuse into a regular $SO(3)$ representation, and $\gamma_g^\alpha(a) = 1, \forall a \in G$.

According to Theorem-I, at least one SRE liquid phase can exist and must be an SPT phase in which the g-symmetry-defect carries a half-integer spin. To understand how many SPT phases are possibly realized, one can follow the Remark. [1] The result is that among all possible SPT phases classified by $H^3(SO(3) \times Z_2^{Ising}, U(1)) = Z \times Z_2^2$, only one of the Z_2 indices is enforced to be nontrivial. And there are many distinct SPT phases that can be realized, which form a coset $\omega_0 \cdot \mathcal{A}_g$, where $\mathcal{A}_g = Z \times Z_2$. *In particular, after gauging the Z_2^{Ising} symmetry, one may obtain either the toric-code or double-semion topological order, depending on which SPT phase is realized.*

(2) $\mathbf{G = Z_2^T \times Z_2^{Ising}}$, and a Kramer doublet per unit cell: Here $Z_2^T = \{I, \mathcal{T}\}$ is the time-reversal symmetrg group. Denoting the Ising and time-reversal transformations on the physical d.o.f. in one unit cell as U_g, and $U_\mathcal{T}$ (antiunitary), the projective representation

[1] SPT phases protected by $G = SO(3) \times Z_2^{Ising}$ form a group $H^3(SO(3) \times Z_2^{Ising}, U(1))$. The Kunneth formula gives: $H^3(SO(3) \times Z_2^{Ising}, U(1)) = H^3(SO(3), U(1)) \times H^3(Z_2, U(1)) \times H^2(SO(3), Z_2) = Z \times Z_2 \times Z_2$. Following the Remark, it is straightforward to show that only the Z_2 index in $H^2(SO(3), Z_2)$ is enforced to be nontrivial. (Namely ω_0 in Theorem-(2) can be chosen to be the nontrivial element in $H^2(SO(3), Z_2)$,and the kernal $\mathcal{A}_g = H^3(SO(3), U(1)) \times H^2(SO(3), Z_2) = Z \times Z_2$.)

α per unit cell satisfies:

$$U_g^2 = 1, \qquad U_{\mathcal{T}} U_{\mathcal{T}}^* = -1, \qquad U_{\mathcal{T}} U_g^* = U_g U_{\mathcal{T}}. \tag{2.18}$$

For instance, this algebra is satisfied if $U_g = \sigma_x$ and $U_{\mathcal{T}} = i\tau_y$ for a four-dimensional local Hilbert space (upon which σ and τ Pauli matrices act). One can check that the two conditions in Theorem-(1) are both satisfied, and thus *at least an SRE liquid phase can exist and must be an SPT phase in which the g-symmetry-defect carries the projective representation α (a Kramer-doublet) above.*

Naively this example is very similar to the example-(1). However there is an important difference. In this example, *only one SPT phase can be realized* — following the Remark, this is because the kernel subgroup $\mathcal{A}_g$ is the trivial Z_1 group. [2] *After gauging the Z_2^{Ising} symmetry, one must obtain a toric code topological order.* This realizable SPT phase is topologically identical to the one obtained by decorating Ising domain walls with the Z_2^T Haldane chains [30].

(3) $\mathbf{G = Z_2^T \times Z_2^{Ising}}$, and a non-Kramer doublet per unit cell: Here the projective representation α per unit cell satisfies:

$$U_g^2 = 1, \qquad U_{\mathcal{T}} U_{\mathcal{T}}^* = 1, \qquad U_{\mathcal{T}} U_g^* = -U_g U_{\mathcal{T}}. \tag{2.19}$$

For instance, $U_g = \sigma_x$ and $U_{\mathcal{T}} = \sigma_z$ on a two-dimensional local Hilbert space would satisfy this algebra. One can check that the two conditions in Theorem-(1) are both satisfied,

[2] Following the Kunneth formula: $H^3(Z_2^T \times Z_2^{Ising}, U(1)) = H^3(Z_2^T, U(1)) \times H^3(Z_2, U(1)) \times H^2(Z_2^T, Z_2) = Z_1 \times Z_2 \times Z_2$. In this example, the Z_2 index in $H^2(Z_2^T, Z_2)$ is enforced to be nontrivial, and the Z_2 index in $H^3(Z_2, U(1))$ is enforced to be trivial. This is because here ω_0 in Theorem-2 is the nontrivial element in $H^2(Z_2^T, Z_2)$, and the kernel subgroup $\mathcal{A}_g = Z_1$.

and thus *at least an SRE liquid phase can exist and must be an SPT phase in which the g-symmetry-defect carries the projective representation α (a Kramer-doublet) above.*

Similar to example-(2), *there is only one realizable SPT phase.* This SPT phase turns out to be the non-trivial Levin-Gu SPT phase protected by the Z_2^{Ising} alone. [3] Namely, here *after gauging the Z_2^{Ising} symmetry, one must obtain a double-semion topological order.*

(4) $\mathbf{G = Z_2' \times Z_2^{Ising}}$, and a projective representation per unit cell: In this example, $Z_2' = \{I, h\}$ is another unitary Ising symmetry group. The projective representation α satisfies:

$$U_g^2 = 1, \qquad U_h^2 = 1, \qquad U_g U_h = -U_h U_g. \tag{2.20}$$

For instance, $U_g = \sigma_x$, $U_h = \sigma_z$ realize this algebra. Two of such projective representations fuse into a regular representation of G, so the condition-(1) in Theorem-(1) is satisified. But one can show that the condition-(2) is *not* satisfied:

$$\gamma_g^\alpha(h) = -1, \tag{2.21}$$

i.e., γ_g^α is a nontrivial 1-cocycle. Therefore according to Theorem-(1), *a SRE liquid phase is not possible.* Without breaking symmetry, this suggests that topological order is inevitable for gapped systems. This is a somewhat surprising result. If one views the system using the enlarged magnetic unit cell, there is no reason why a SRE liquid is not allowed.

[3] Now in this example, ω_0 in Theorem-2 is the nontrivial element in $H^3(Z_2, U(1))$, and the kernel subgroup $\mathcal{A}_g = Z_1$.

2.4 Decorated Quantum Dimer Models for SPT phases

Closely related to the decorated-BFG model in Sec.2.2, in this section we describe a class of exactly solvable models realizing symmetry-enforced SPT phases. These models are constructed by decorating quantum dimer models(QDM) with relevant physical degrees of freedom, whose ground states can be exactly solved at the corresponding Rokhsar-Kivelson point [27]. Although this class of models can be generalized to other lattices, here we will focus on the decoration of the QDM on the triangular lattice [32, 33]. In particular, we will construct models realizing the symmetry-enforced SPT phases in example-(1,2,3) in Sec.2.3.1

2.4.1 $G = SO(3) \times Z_2^{Ising}$, a spin-1/2 per unit cell

Continuing with discussions in example-(1) in Sec.2.3.1, in the presence of onsite global symmetry $G = SO(3) \times Z_2^{Ising}$, we consider quantum systems with one spin-1/2 per unit cell in two spatial dimensions respecting the Ising magnetic translation symmetry Eq.(2.16). Note that we will reserve symbols T_x, T_y for the magnetic translations, and use $T_x^{orig.}, T_y^{orig.}$ to represent the original translations.

We will construct two exactly solvable models (model-A and model-B) respecting the symmetry described above featuring SRE liquid ground states. Although the Ising defects in both models carry half-integer spins, the two models are in distinct SPT phases. The simplest way to understand their difference is that, after gauging the Ising symmetry, model-A has toric-code topological order while model-B has double-semion topological order.

We start with constructing model-A. This model contains two sets of degrees of freedom (d.o.f.): the Ising-d.o.f. σ_I which live on a honeycomb lattice and the spin-1/2-d.o.f. $\mathbf{S}_i =$

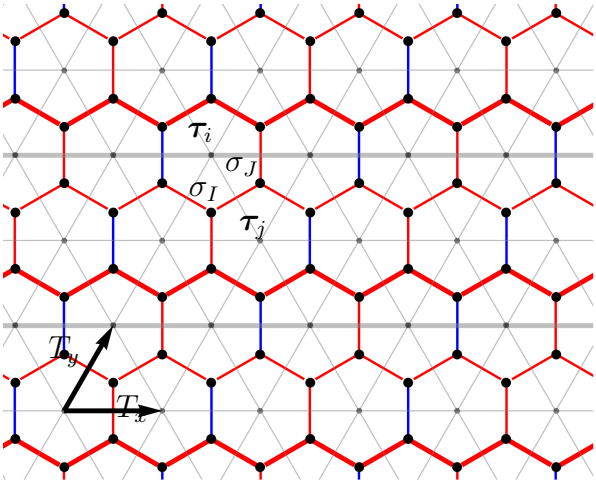


Figure 2.4: (color online) The Ising d.o.f. σ live on the honeycomb lattice and the spin d.o.f. $\boldsymbol{\tau}$ lives on the triangular lattice. The Ising coupling signs $s_{IJ} = +1$ on red bonds, and $s_{IJ} = -1$ on blue vertical bonds. The thick red bonds represent the "y-odd zigzag chains" used in Eq.(2.25,2.36). The thick gray horizontal bonds on the triangular lattice represent the "y-odd rows" used in Eq.(2.36).

$\boldsymbol{\tau}_i/2$ which live on the triangular lattice formed by centers of the hexagons, as shown in Fig.2.4. The Hamiltonian of model-A contains three terms:

$$H = H^{Ising} + H^{binding} + H^{A}. \tag{2.22}$$

where H^{Ising} is simply a frustrated nearest-neighbor Ising model:

$$H^{Ising} = -K \sum_{\langle IJ \rangle} s_{IJ} \sigma_I^z \sigma_J^z. \tag{2.23}$$

Here σ_I^z, σ_J^z are the Ising spins living on the honeycomb sites labled by I, J, the coupling constant $K > 0$, and $s_{IJ} = \pm 1$ defined as in Fig.2.4. $H^{binding}$ is an interaction between the Ising-d.o.f. and spin-1/2-d.o.f., which commutes with H^{Ising}:

$$H^{binding} = -\lambda \sum_{\substack{\text{J} \\ \text{i} + \text{j} \\ \text{I}}} \frac{1}{2}(1 - s_{IJ}\sigma_I^z \sigma_J^z) \cdot \hat{P}_{\mathbf{S}_i + \mathbf{S}_j = 0}, \tag{2.24}$$

where $\lambda > 0$, i, j labels the sites on the triangular lattice, and the summation of "+" is over all intersection points between the triangular lattice and the honeycomb lattice as shown in Fig.2.4. $\hat{P}_{\mathbf{S}_i + \mathbf{S}_j = 0} \equiv \frac{1}{4} - \mathbf{S}_i \cdot \mathbf{S}_j$ is the operator projecting the two spin-1/2's on site-i and site-j into a spin singlet.

H^A is more complicated and will be given in Eq.(2.29). It is straightforward to checked that H respects the Ising symmetry $U_g \equiv \prod_I \sigma_I^x$, the spin-rotation symmetry generated by $\sum_i \mathbf{S}_i$. H also respects magnetic translation operations:

$$T_x = \Big(\prod_{I \in \text{y-odd zigzag chains}} \sigma_I^x \Big) \cdot T_x^{orig.}, \qquad T_y = T_y^{orig.}, \tag{2.25}$$

(see Fig.2.4), and the magnetic translation algebra Eq.(2.16) is satisfied. We will show that the ground state of H is in a gapped liquid phase without topological order. According

to our general results, this ground state must be in an SPT phase, which we will show momentarily.

The physical consequence of H^{Ising} and $H^{binding}$ is to provide a highly degenerate low energy manifold, which will be lifted by H^A. To understand the low energy manifold, let us firstly consider H^{Ising}. Because every hexagonal plaquette frustrated, there will be at least one bond in each plaquette that is energetically unhappy. Namely the ground state manifold of H^{Ising} is formed by all possible Ising configurations satisfying the "one-unhappy-Ising-bond-per-plaquette" condition.

$H^{binding}$ further constrains the spin-1/2 d.o.f. in the low energy manifold. It has effect only on the Ising unhappy bonds ($s_{IJ}\sigma^z_I\sigma^z_J = -1$), and energetically binds the two spin-1/2's near the unhappy Ising-bond into a spin singlet. The degenerate ground state manifold of $H^{Ising} + H^{binding}$ is now clear: it is formed by all such quantum states satisfying the "one-unhappy-Ising-bond-per-plaquette" condition and the two neighboring spin-1/2's normal to every unhappy Ising-bond form a spin singlet.

It is well-known that the Ising configurations satisfying the "one-unhappy-Ising-bond-per-plaquette" condition are intimately related to the Hilbert space of the QDM [34–36]. Pictorically, any such an Ising configuration can be mapped to a dimer covering (with one dimer per site) on the triangular lattice by assigning a dimer crossing the unhappy Ising bond. The effect of $H^{binding}$ is simply to energetically binds the two spin-1/2's in each dimer into a spin-singlet. Namely if the Ising configuration is given, the state of spin-1/2 d.o.f. is also fixed.

Similar to the model in Sec.2.2, there is a two-to-one mapping from the ground state manifold of $H^{Ising} + H^{binding}$ to the QDM Hilbert space, since two low energy states differ

by a global Ising transformation map into the same state in the QDM. Second, these states only map to one specific topological sector of the QDM: The parity of the number of dimers crossing a loop is simply given by the sign of the product $\prod_{\langle IJ\rangle \in loop} s_{IJ}$.

In fact, these relations between the Ising d.o.f. and the dimer d.o.f. can be viewed as the well-known duality mapping between quantum Ising models and Z_2 gauge theories [37]. The Ising paramagnet phase is dual to the deconfined Z_2 gauge phase in the QDM, while the Ising ordered phase is dual to the confined Z_2 gauge theory. More precisely, the "unhappy-Ising-bond" is dual to the electric flux line (i.e., the dimer) in the gauge theory, and the "one-unhappy-Ising-bond-per-plaquette" condition in the low energy manifold on the Ising side is dual to the "one-Z_2-gauge-charge-per-site" condition in the QDM. Here, the only new ingredient apart from this well-known duality is that there are also spin-1/2 d.o.f. on the Ising side. But due to $H^{binding}$, these spin-1/2's form a pattern of spin singlets fixed by the Ising d.o.f., and consequently the duality mapping is not modified after a convention of the spin-singlet signs is given (see below).

After introducing H^A, we will see that the degeneracy in the low energy manifold will be lifted and the unique ground state on a torus sample is formed. The usual QDM Hamiltonian on the triangular lattice is:

$$H^{TC}_{QDM} = -t \sum_{\text{plaquettes}} \big(|\blacklozenge\rangle\langle\square| + h.c. \big) + v \sum_{\text{plaquettes}} \big(|\blacklozenge\rangle\langle\blacklozenge| + |\square\rangle\langle\square| \big), \tag{2.26}$$

where the summation is over all plaquettes (rhombi): "$\square$", "$\square$", "$\diamondsuit$". The ground states of this model are exactly known at the RK-point given by $t = v > 0$, and the superscript TC is highlighting that the topological order is toric-code [32, 38] like in the deconfined phase

(i.e., the usual Z_2 gauge theory). At this point H^{TC}_{RK} can be rewritten as a summation of projectors [32]:

$$H^{TC}_{RK} = t \sum_{\text{plaquettes}} \big(|\square\rangle - |\square\rangle\big)\big(\langle\square| - \langle\square|\big). \tag{2.27}$$

Clearly the equal weight superposition of all dimer coverings within any fixed topological sector $|\Phi^{TC}_{RK}\rangle = \sum_c |c\rangle$ (c labels possible dimer coverings) is one ground state of H^{TC}_{RK} since it is annihilated by all projectors.

Based on the duality mapping, H^{TC}_{QDM} is mapped to H^A. Any dimer covering $|c\rangle$ will be mapped to two Ising configurations $|c_1\rangle$ and $|c_2\rangle$ distinct from each other by a global Ising transformation. One can further choose a translationally symmetric sign convention for the spin-singlets on the nearest neighbor bonds along the three orientations:

$$\begin{aligned}
|\text{—}\rangle &\equiv \frac{1}{\sqrt{2}}\big(|\uparrow\!\!-\!\!\downarrow\rangle - |\downarrow\!\!-\!\!\uparrow\rangle\big), \\
|\ \diagup\ \rangle &\equiv \frac{1}{\sqrt{2}}\big(|\ \downarrow\!\diagdown\!\uparrow\ \rangle - |\ \uparrow\!\diagdown\!\downarrow\ \rangle\big), \\
|\ \diagup\ \rangle &\equiv \frac{1}{\sqrt{2}}\big(|\ \downarrow\!\diagup\!\uparrow\ \rangle - |\ \uparrow\!\diagup\!\downarrow\ \rangle\big)\rangle.
\end{aligned} \tag{2.28}$$

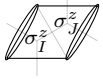
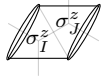
With this sign convention, given a c, $|c_i\rangle$ ($i = 1$ or 2) fully determines a state in the ground state manifold of $H^{Ising} + H^{binding}$ by replacing the dimer configuration by the corresponding spin-singlet configuration. In addition, $\{|c_1\rangle, |c_2\rangle\}$ for all c form a complete orthornormal basis in this manifold.

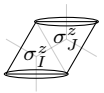
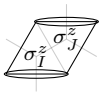
H^A is defined as:

$$H^A = -t \sum_{\text{plaquettes}}^{\sigma_I^z,\sigma_J^z=\pm 1} \left(\left| \right. \right.$$

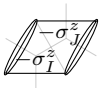
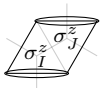
$$\left. \left\rangle\right\langle \right.$$

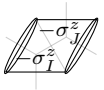
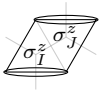
$$\left. \left| + h.c. \right. \right)$$

$$+ v \sum_{\text{plaquettes}}^{\sigma_I^z,\sigma_J^z=\pm 1} \left(\left| \right. \right.$$

$$\left. \left\rangle\right\langle \right.$$

$$\left. \left| \right. \right.$$

$$\left. + \left|\right. \right.$$

$$\left. \left\rangle\right\langle \right.$$

$$\left. \left| \right. \right) \tag{2.29}$$

Note that the t-term also flips the two Ising spins inside the plaquette. At the RK point $t = v > 0$, H^A can again be written as a summation of projectors:

$$H^A_{RK} = t \sum_{\text{plaquettes}}^{\sigma_I^z,\sigma_J^z=\pm 1} \left(\left| \right. \right.$$

$$\left. \left\rangle - \right| \right.$$

$$\left. \left\rangle \right. \right)$$

$$\cdot \left(\left\langle \right. \right.$$

$$\left. \left| - \right\langle \right.$$

$$\left. \left| \right. \right) \tag{2.30}$$

To study the ground state of the total Hamiltonian H, it is suffice to focus on the degenerate ground state manifold of $H^{Ising} + H^{binding}$, and clearly H^A acts within this manifold. In addition, it is straightforward to show that $|\Phi^A_{RK}\rangle = \sum_c (|c_1\rangle + |c_2\rangle)$, i.e, the equal weight superposition of all states in this manifold, is a ground state of H^A_{RK} because it is annihilated by every projector in Eq.(2.30). $|\Phi^A_{RK}\rangle$ is clearly a fully symmetric liquid wavefunction.

It is known that for the QDM Eq.(2.26), the RK point of H is exactly at a first-order phase transition boundary between a deconfined gapped liquid phase ($v < t$) and a staggered valence bond solid phase ($v > t$) [32]. Based on the duality mapping, the model-A is in a fully symmetric gapped liquid phase for $v_c < v < t$ with a unique ground state on torus. In the limit of $K, \lambda \gg v, t$, v_c is given by the same critical value $v_c \approx 0.7t$ as in the original QDM [32]. More precisely, in the global Ising-even sector of the Hilbert space of model-A, this mapping to the Hilbert space of the QDM is given by $1/\sqrt{2}(|c_1\rangle + |c_2\rangle) \to |c\rangle$, and clearly H^A is mapped to H^{TC}_{QDM}. Namely, the full energy spectrum of model-A in the

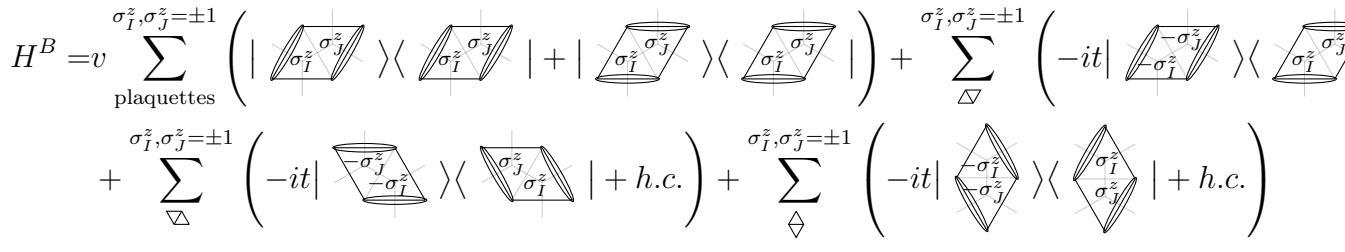
Ising-even sector has a one-to-one correspondence with the full energy spectrum of H^{TC}_{QDM}. In addition, in the Ising paramagnetic phase, ground state in the global Ising-odd sector of model-A has a finite excitation energy which is the same as the energy cost of a Z_2 gauge flux in the QDM.

Next we show that the liquid phase $v_c < v < t$ in model-A is an SPT phase because the Ising defects carry half-integer spins. Similar to the discussion in Sec.2.2 (see Fig.2.3), after a pair of Ising defects are spatially separated the original Hamiltonian H is modified into H'. Comparing with H, the s_{IJ} flips sign in H' whenever the bond $I-J$ crosses the branch cut. Namely, for any loop on the honeycomb lattice enclosing a single Ising defect, the product $\prod s_{IJ}$ along the loop changes sign. In order not to cost $H'^{binding}$ energy, the parity of the number of dimers crossing this loop also flips. Consequently, an Ising defect is topologically bound with a monomer (an unpaired site on the triangular lattice). This monomer clearly carries a half-integer spin in model-A, which can be detected by the local spin susceptibility at low temperatures.

Next, we demonstrate a different symmetry-enforced SPT phase using the model-B defined as follows:

$$H = H^{Ising} + H^{binding} + H^{B}. \tag{2.31}$$

Comparing with the model-A in Eq.(2.22), only the last term is modified:

$H^B =$ 

(2.32)

One can straightforwardly check that the model-B defined in Eq.(2.31,2.32) also respects the Ising symmetry, the $SO(3)$ spin-rotation symmetry, and the magnetic translation symmetry Eq.(2.16). Below we show that in a finite parameter regime $v_c' < v < t$, the model-B is in a gapped liquid phase, and this phase is another SPT phase.

In the ground state manifold of $H^{Ising} + H^{binding}$, the duality transformation maps H^B into the following QDM Hamiltonian:

$$H^{DS}_{QDM} = v \sum_{\text{plaquettes}} \big(|\square\rangle\langle\square| + |\square\rangle\langle\square| \big)$$
$$+ \sum_{\square,\square,\diamondsuit} -it\big(|\square\rangle\langle\square| + |\square\rangle\langle\square| + |\diamondsuit\rangle\langle\diamondsuit| \big) + h.c. \tag{2.33}$$

This QDM was firstly introduced and studied in Ref. [33], where the exactly solvable RK point $t = v$ has been shown to be adjacent a gapped liquid phase for $v \lesssim t$. Interestingly, this phase was demonstrated to have a double-semion topological order (the superscript DS here is to highlight this fact). By the duality mapping, we know that in a finite parameter regime $v_c' < v < t$, the model-B is in a gapped liquid phase.

Similar to previous disussion on the model-A, it is straightforward to show that the Ising defect in the gapped liquid phase of model-B also carries half-integer spin, so it is also an SPT phase. To see the difference from the SPT phase realized in model-A, let us consider the Ising symmetry only. It is known that Ising symmetry itself can protect two paramagetic phases: the trivial phase and the SPT phase. Levin and Gu pointed out [39] that the duality mapping maps the usual Ising paramagnet to the toric-code topological order, while the nontrivial Ising SPT phase maps to the double semion topological order. Consequently, the SPT phases realized in model-A and model-B are different because the Ising symmetry alone already distinguishes them.

2.4.2 $G = Z_2^T \times Z_2^{Ising}$

Here we demonstrate symmetry-enforced SPT phases outlined in example-(2) and (3) in Sec.2.3.1. Unlike the $G = SO(3) \times Z_2^{Ising}$ case, here we show that symmetry conditions fully determine the SPT phase. The models below have the same Hilbert space as in the $G = SO(3) \times Z_2^{Ising}$ case (i.e., σ_I on the honeycomb lattice and $\boldsymbol{\tau}_i$ on the triangular lattice), but with different symmetries defined.

A Kramer doublet per unit cell: A simple generalization is for example-(2) (i.e., a Kramer-doublet per unit cell) where we can recycle the model-A. Namely, defining the Ising symmetry $U_g = \prod_I \sigma_I^x$ as before and the antiunitary time-reversal symmetry as $U_{\mathcal{T}} = e^{i\pi \mathbf{S}^y} = i\boldsymbol{\tau}^y$, clearly model-A respect all the required symmetries. In addition, we have one Kramer doublet $\boldsymbol{\tau}$ per unit cell. According to our discussion in example-(2), this SRE liquid phase realized in $v_c < v < t$ must be an SPT phase in which the Ising defect carries a Kramer doublet, which is obviously realized in model-A. In addition, after gauging Z_2^{Ising} symmetry, one necessarily obtains the toric-code topological order, which is confirmed in model-A.

On the other hand, model-B, gauging which gives double-semion topological order, explicitly breaks the time-reversal symmetry defined above.

A non-Kramer doublet per unit cell: Now let us move on to example-(3) in Sec.2.3.1. In order to construct a model realizing the non-Kramer doublet projective representation defined in Eq.(2.19), let us define the following symmetry operations:

$$U_g = \boldsymbol{\tau}^z \prod_I \sigma_I^x, \qquad\qquad U_{\mathcal{T}} = \boldsymbol{\tau}^x. \tag{2.34}$$

Consequently we have one non-Kramer doublet $\boldsymbol{\tau}$ per unit cell. The model Hamiltonian in

this example will be given by

$$H = H^{Ising} + \tilde{H}^{binding} + \tilde{H}^{B}, \tag{2.35}$$

where H^{Ising} is given in Eq.(2.23), $\tilde{H}^{binding}$ is in Eq.(2.39). $\tilde{H}^{B}$ has the same form as H^{B} in Eq.(2.32), but with a modified interpretations of the dimers: replacing the $|\rangle$'s defined in Eq.(2.28) by $|\rangle$'s and $|\rangle$'s defined in Eq.(2.37,2.38) depending the dimer positions (see discussions below and Fig.2.5). Eventually we will show that in a finite regime $v \lesssim t$ this model features a SRE gapped liquid ground state which is the Levin-Gu Ising SPT phase, consistent with discussions in Sec.2.3.1.

We will construct a model similar to $H = H^{Ising} + H^{binding} + ...$, but respecting a magnetic translation operations different from Eq.(2.25) since U_g is now different. In particular, we define magnetic translations:

$$\begin{aligned} T_x &= \Big(\prod_{i \in \text{y-odd row}} \boldsymbol{\tau}_i^z\Big)\Big(\prod_{I \in \text{y-odd zigzag chains}} \sigma_I^x\Big) T_x^{orig.}, \\ T_y &= T_y^{orig.}, \end{aligned} \tag{2.36}$$

(see Fig.2.4), which satisifies Eq.(2.16) with $g = \prod_i \boldsymbol{\tau}_i^z \prod_I \sigma_I^x$ following Eq.(2.34). Although H^{Ising} still respects all the symmetries, $H^{binding}$ does not respect g and T_x. This is because the usual spin singlets $|\rangle$ in the projector $\hat{P}_{\mathbf{S}_i+\mathbf{S}_j=0} = |\rangle\langle|$ does not respect g and T_x. We therefore need to modify these dimer states and the projectors.

We define the following dimer states formed by the $\boldsymbol{\tau}$ spins:

$$\begin{aligned} |\rangle &\equiv \frac{1}{\sqrt{2}}\big(|\uparrow\!\!-\!\!\uparrow\rangle + |\downarrow\!\!-\!\!\downarrow\rangle\big), \\ |\ \ \rangle &\equiv \frac{1}{\sqrt{2}}\big(|\ \ \rangle + |\ \ \rangle\big), \\ |\ \ \rangle &\equiv \frac{1}{\sqrt{2}}\big(|\ \ \rangle + |\ \ \rangle\big)\rangle. \end{aligned} \tag{2.37}$$

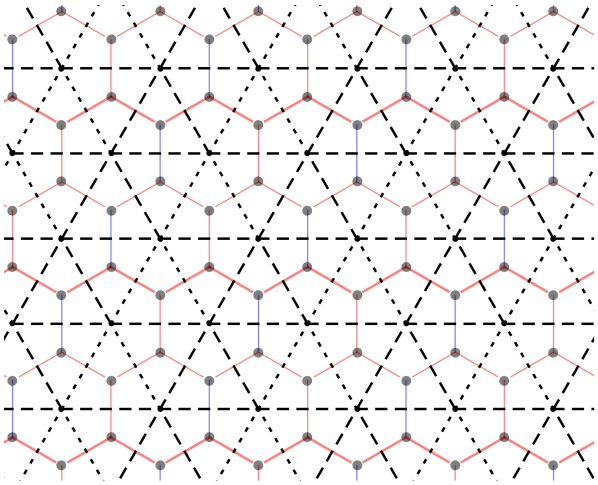


Figure 2.5: To construct model in Eq.(2.35), the dimer states living on the nearest neighbor bonds on the triangular lattice have a spatial dependent pattern: the dimer states living on the dashed bonds are defined in Eq.(2.37), and those living on the dotted bonds are defined in Eq.(2.38).

Clearly these dimer states are both Ising and time-reversal even according to Eq.(2.34).

Under the magnetic translation T_x in Eq.(2.36), dimer states with $|$ ⟍ $\rangle$ and $|$ ⫽ $\rangle$ connecting even and odd rows on the triangular lattice will transform into the following states:

$$|\ \diagdown\ \rangle \equiv \frac{1}{\sqrt{2}}(|\ \uparrow\uparrow\ \rangle - |\ \downarrow\downarrow\ \rangle),$$
$$|\ /\!/\ \rangle \equiv \frac{1}{\sqrt{2}}(|\ \uparrow\uparrow\ \rangle - |\ \downarrow\downarrow\ \rangle)\rangle. \tag{2.38}$$

Note that these dotted dimer states are Ising even, but *time-reversal odd* according to Eq.(2.34). The magnetic translational symmetric assignment of the dimer states is given in Fig.(2.5): the states $|$ ⟍ $\rangle$, $|$ ⫽ $\rangle$ are assigned on the dashed bonds, while the states $|$ ⟍ $\rangle$, $|$ ⫽ $\rangle$ are assigned on the dotted bonds.

By replacing the spin singlets in $\hat{P}_{\mathbf{S}_i+\mathbf{S}_j=0} = |\text{⬭}\rangle\langle\text{⬭}|$ by the corresponding $|$ ⟍⫽ $\rangle\langle$ ⟍⫽ $|$

and $|$$\rangle\langle$$|$ in the spatial dependent fashion in Fig.2.5, we modify $H^{binding}$ naturally as

$$\tilde{H}^{binding} = -\lambda \sum_{\substack{\text{J} \\ \text{i} \text{ -}\!\!+\!\!\text{- j} \\ \text{I}}} \frac{1}{2}(1 - s_{IJ}\sigma_I^z\sigma_J^z) \cdot \left(|\text{}\rangle\langle\text{}|\right)$$

$$- \lambda \sum_{\substack{\text{J} \\ \text{i} \cdots\!\!+\!\!\cdots \text{ j} \\ \text{I}}} \frac{1}{2}(1 - s_{IJ}\sigma_I^z\sigma_J^z) \cdot \left(|\text{}\rangle\langle\text{}|\right), \tag{2.39}$$

Now $H^{Ising} + \tilde{H}^{binding}$ respects all the required symmetries, but one still need terms like H^A or H^B to lift the ground state degeneracy to reach a SRE gapped liquid phase. In order to preserve the Ising and magnetic translation symmetry, it is natural to replace the dimer states $|$$\rangle$ by the corresponding $|$$\rangle$ and $|$$\rangle$. Let us denote the resulting modified Hamiltonians as $\tilde{H}^A$ or $\tilde{H}^B$. Both modified models $H^{Ising} + \tilde{H}^{binding} + \tilde{H}^A$ and $H^{Ising} + \tilde{H}^{binding} + \tilde{H}^B$ are solvable, by duality mapping to the H^{TC}_{QDM} and H^{DS}_{QDM} respectively. These two models both give SRE gapped ground states in the regime $v \lesssim t$ respecting the magnetic translation symmetry and the Ising symmetry, gauging which give toric-code and double semion topological order respectively.

Finally, let us consider the time-reversal symmetry in Eq.(2.19). Importantly, $|$$\rangle$'s are time-reversal even while $|$$\rangle$ are time-reversal odd. As shown in the pattern Fig.2.5, any dimer resonant term like $|$$\rangle\langle$$|$ will involve an odd number of $|$$\rangle$ states. Consequently, *only $\tilde{H}^B$ is time-reversal symmetric, while $\tilde{H}^A$ explicitly breaks the time-reversal.* In fact according to Theorem-I, it is impossible to have a SRE liquid respecting all the required symmetries and gauging the Ising symmetry gives a toric code topological order.

2.5 Proof of Theorems

Here we present a combination of physical argument and mathematical derivations based on symmetric tensor network formulation [21, 40–47]. We will focus on Theorem-II, and in Appendix2.7.2 we show that Theorem-I can be viewed as its special case.

We need to show the condition in Theorem-II is necessary and sufficient for a SRE liquid phase to exist, which must be an SPT phase. To show this condition is necessary, we consider such a SRE liquid phase and the pumping of the entanglement spectra during an adiabatic process in Sec.2.5.1, leading to an observation that a g-symmetry defect must carry a projective representation α. In a SRE liquid phase characterized by 3-cocycle ω, we use the symmetric tensor-network formulation in Sec.2.5.2 to establish that the projective representation carried by a g-symmetry defect is given by $(\delta_g^\omega)^{-1}$. Since the projective representation carried by a symmetry defect is physical and independent of formulation, together with the pumping argument, the necessary condition in Theorem-II is established *independent of formulation*.

As a complementary calculation, we also explicitly compute the projective representation carried by a g-symmetry defect in a SRE liquid phase representable within the symmetry tensor-network formulation in Appendix.2.7.5, which turns out to be α, consistent with the previous discussion [48].

To further show that the condition is also sufficient, we will show that for any 3-cocycle ω_0 satisfying $(\delta_g^{\omega_0})^{-1} \simeq \alpha$, generic symmetric tensor network wavefunctions representing a SRE liquid phase characterized by ω_0 can be constructed in Sec.2.5.3.

2.5.1 Entanglement Pumping argument

Here let us assume a SRE liquid phase exist. It is straightforward to show that after a local unitary transformation (a site dependent U_g action) without changing the physical action of G, one can always choose a gauge in which $T_y = T_y^{orig.}$ and $T_x = (\prod_{y-odd} U_g) \cdot T_x^{orig.}$. Then we consider putting this SRE liquid on a infinite cylinder C along the x-direction, with L_y number of unit cells across the y-direction loop. We choose $L_y = l_y \cdot N + d_y$ where l_y, d_y are integers, $g^N = I$ and $0 \leq d_y < N$. Note that if $d_y \neq 0$ this choice of L_y will explicitly break the T_x symmetry.

This infinite long cylinder can be viewed as a one-dimensional system, and for a large enough L_y the onsite symmetry G will be respected. We will study the entanglement spectrum of this one dimensional system at a particular cut $x_0 + 1/2$. A SRE liquid respecting G symmetry dictates that entanglement states at this cut carry a particular projective representation ξ of G [49, 50].

Next, we adiabatically create a g-symmetry defect/anti-defect pair at a given y-coordinate and separate them to infinity along the x-direction. After repeating this procedure for every y-coordinate, we totally move L_y number of g-symmetry defects acrosses the entanglement cut $x_0 + 1/2$. As shown in Fig.2.6, the final Hamiltonian is related with the original Hamiltonian by the *original* translation operation $T_x^{orig.}$. Therefore, the final entanglement states at the cut $x_0 + 1/2$ is equivalent to the initial entanglement states at a different cut: $x_0 - 1/2$.

Because the initial entanglement states at $x_0 - 1/2$ differs from the initial entanglement states at $x_0 + 1/2$ by a column of unit cells along y-direction, we conclude that the final entanglement eigenstates at $x_0 + 1/2$ must carry the $\alpha^{d_y} \cdot \xi$ projective representation. This pumping of the entanglement projective representation can only be explained by the pro-

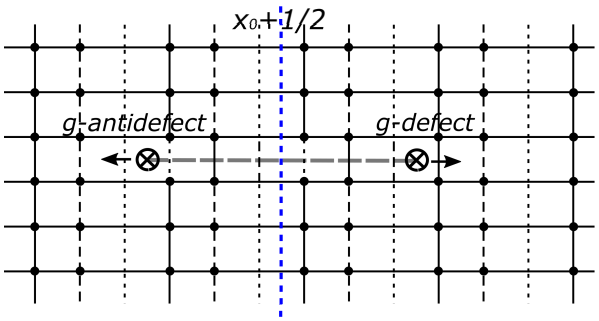


Figure 2.6: Illustration of adiabatically separating a pair of g-defect/antidefect along the x-direction with $g^3 = I$. For simplicity, one may imagine Hamiltonian to host nearest neighbor (NN) terms. Along the x-direction, due to the magnetic translation symmetry Eq.(2.16), the NN interactions on the vertical bonds have a three-unit-cell periodicity (solid,dashed and dotted bonds). While the g-defect crosses the entanglement cut at $x_0 + 1/2$, the Hamiltonian along the branch cut (dashed gray line) is effectively translated along x-direction by one unit cell. After separating such pairs of defects for every row, the final Hamiltonian is related to the original Hamiltonian by $T_x^{orig.}$.

jective representation $\tilde{\alpha}$ carried by a g-symmetry defect, and $(\tilde{\alpha})^{L_y} \simeq \alpha^{d_y}$. But because N g-defect fuse into a trivial object, we must have $(\tilde{\alpha})^N \simeq 1$. Consequently $\tilde{\alpha} = \alpha$.

2.5.2 Symmetry-enforced constraints on SPT cocycles

Here we consider an SPT wavefunction represented using the symmetric tensor-network formulation [21]. The advantage of this formulation is that it allows us to introduce symmetry defects conveniently.

The local symmetry transformation of an onsite symmetry a on a g-defect is given by the application of of U_a inside a disk D covering the g-defect, together with an application of unitary operations on the virtual degrees of freedom on the boundary of D. This boundary operation should be defined in such a way that after these two operations, no excitation is created near the boundary of D. In Appendix 2.7.4 we explictly constructed such boundary operations. With these boundary operations, we explicitly show that the projective representaion carried by the g-defect is given by $(\delta_g^\omega)^{-1}$ in an SPT phase characterized by the 3-cocycle ω, which is given without proof in Ref. [48].

2.5.3 Generic constructions of Symmetry-enforced SPT wavefunctions

Our strategy here is to start from an SPT state characterized by a 3-cocycle ω with a regular representation of G per unit cell and respecting the usual translation symmetry $T_x^{orig.}, T_y^{orig.}$. Such a state can be generically represented using the symmetric tensor-network formulation [21]. In particular, the symmetric tensor-network needs to satisfy a collection of algebraic equations (constraints). Then we show that after properly modifying these alge-

braic constraints, the new tensor-network will respect the magnetic translation symmetry T_x, T_y, at the same time the physical d.o.f. must carry a projective representation $(\delta_g^\omega)^{-1}$ per unit cell (otherwise the wavefunction vanishes). The tensor-network states satisfying these modified constraints are generic constructions of the SPT states in Theorem-II.

The details of the construction can be found in Appendix 2.7.6.

2.6 Discussion

Generalized Hastings-Oshikawa-Lieb-Schultz-Mattis theorems put strong constraints on possible symmetric quantum states of matter. In particular, in the presence of translation symmetry and a projective representation of the onsite symmetry group per unit cell, it is impossible to have a gapped short-range entangled (SRE) symmetric ground state. In this chapter we discuss that in the presence of magnetic translation symmetry, gapped SRE symmetric ground states could exist, which are enforced to be symmetry protected topological (SPT) phases. Focusing on bosonic systems in two spatial dimensions, we provide the generic necessary and sufficient condition for such symmetry-enforced SPT phases to occur in Theorem-I and II. When the condition is satisified, we sharply characterize the coset structure of the realizable SPT phases in the Remark.

The condition-(2) in Theorem-I is particular non-obvious. It states that if symmetries protecting the fractional spin (projective representation) per unit cell and those generating the magnetic translations fail to commute with one another, then SRE liquid state is impossible even if the fractional spins fuse into an integer spin (regular representation) in the magnetic unit cell.

In addition, we design a class of decorated quantum dimer models realizing some of these symmetry-enforced SPT phases, which are exactly solvable at the corresponding Rokhsar-Kivelson points. A particularly simple model realizing a symmetry-enforced SPT phase is given in Sec.2.2 by coupling a Balents-Fisher-Girvin spin liquid with a layer of pure-transverse-field Ising spins via three-spin interactions. This model also demonstrates the route to obtain SPT phases via condensing anyons in SET phases [21, 22].

It is interesting to consider the situation of fermions with magnetic translation symmetries, in which case (generalized) Hastings-Oshikawa-Lieb-Schultz-Mattis theorem apply for fractional filled systems with regular translation symmetries. In fact, earlier works [13, 14] establish the magnetic translation symmetry protected integer Hall conductivity, and a recent work by Wu et.al. studied the magnetic translation enforced quantum spin Hall insulators in fractionally filled fermionic systems [15]. Connecting with these works, the present work focuses on bosonic systems with projective representation per unit cell, but obtains systematic results.

2.7 Appendices

2.7.1 Perturbation study of the decorated Balents-Fisher-Girvin model

The Hamiltonian for the decorated BFG model can be split into two parts

$$\begin{aligned} H^{deco.BFG} &= H_0 + H_1, \\ H_0 &= J_z \sum_{\hexagon} (S^z_{\hexagon})^2 - \sum_{\mathrm{I} \overset{\mathrm{i}}{\bullet\!\!-\!\!\bullet} \mathrm{J}} \lambda S^z_i (s_{IJ} \sigma^z_I \sigma^z_J), \\ H_1 &= J_\perp \sum_{(i,j)} S^+_i S^-_j + \sum_I h \sigma^x_I, \end{aligned} \tag{2.40}$$

where (i, j) runs over first, second, third neighbors within a hexagon of Kagome plaquette.

Let's take the limit $J_z, \lambda \gg J_\perp, h$ and only focus on the low energy Hamiltonian in the ground state manifold of H_0. Then we can treat H_1 as a small perturbation and use the conventional Brillouin-Wigner perturbation to derive the effective Hamiltonian. The effective Hamiltonian is then given by (take E_0 as the ground state energy of H_0),

$$H_{eff} = E_0 + P_g(H_1 + H_1 G'_0 H_1 + H_1 G'_0 H_1 G'_0 H_1 + \cdots) P_g, \tag{2.41}$$

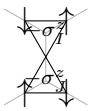
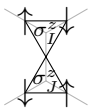
where $G'_0 = P_e(E_0 - H_0)^{-1} P_e$ and P_g/P_e are the projector onto the the ground/excited states of H_0.

We will work under the limit $J_z \gg \lambda$ and calculate the effective Hamiltonian order by order to find the leading non-constant terms in $J_\perp$ and h since we have not specified the relation between them yet. Let's denote N as the number of Kagome unit-cell. The zeroth order energy is $E_0 = -\frac{3}{2} N\lambda$. Higher order terms are as follows:

1. $H^{(1)}_{eff} = P_g H_1 P_g = 0$.

2. $H_{eff}^{(2)} = P_g H_1 G_0' H_1 P_g = -9N\frac{J_\perp^2}{2J_z+2\lambda} - 2N\cdot\frac{h^2}{3\lambda}$. The first term comes from the process of switching a pair of spin-up and spin-down and then switching back within a hexagon. The second term comes from the process of flipping a Ising d.o.f. twice. To this order, we only have constant terms.

3. $H_{eff}^{(3)} = P_g H_1 G_0' H_1 G_0' H_1 P_g = \frac{J_\perp^3}{(2J_z+2\lambda)^2} \sum_{(i,j,k)} (S_i^+ S_k^- S_k^+ S_j^- S_j^+ S_i^- + S_j^+ S_k^- S_i^+ S_j^- S_k^+ S_i^-)$, where the summation runs over all ordered triplets (i,j,k) with $(i,j),(j,k),(k,i)$ appearing in H_1. The term $S_i^+ S_k^- S_k^+ S_j^- S_j^+ S_i^- = (1/2 + S_i^z)(1/2 - S_j^z)(1/2 - S_k^z)$ measures the energy of the configuration with $S_i^z = 1/2, S_j^z = -1/2, S_k^z = -1/2$. And the term $S_j^+ S_k^- S_i^+ S_j^- S_k^+ S_i^- = (1/2 + S_i^z)(1/2 + S_j^z)(1/2 - S_k^z)$ measures the energy of the configuration with $S_i^z = 1/2, S_j^z = 1/2, S_k^z = -1/2$. To this order, the term does depend on the spin configuration and is the leading non-constant term in $J_\perp$.

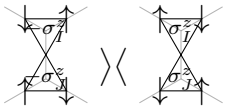
4. $H_{eff}^{(4)} = P_g H_1 G_0' H_1 G_0' H_1 G_0' H_1 P_g = -\frac{10 J_\perp^2 h^2}{9 J_z \lambda^2} \sum_{\bowtie} (|$  $\rangle\langle$  $| + h.c.) + \mathcal{O}(\frac{h^4}{\lambda^3}) + \mathcal{O}(\frac{J_\perp^2 h^2}{J_z \lambda^2}) + \mathcal{O}(\frac{J_\perp^4}{J_z^3})$, where we have used the limit $J_z \gg \lambda$. The first term is a kinetic term which is the leading non-constant contribution in h. The latter three terms are not written out explicitly due to the following reason. Terms proportional to $\frac{J_\perp^4}{J_z^3}$ are less significant compared to that from the 3-rd order perturbation. Terms proportional to $\frac{h^4}{\lambda^3}$ (process of flipping two different Ising d.o.f. twice) is a constant. And the potential term proportional to $\frac{J_\perp^2 h^2}{J_z \lambda^2}$ (process of separately flipping Ising d.o.f twice and exchanging spin-up and down twice) is also a constant in the limit $J_z \gg \lambda$.

The leading non-constant terms are terms of order $\frac{J_\perp^3}{J_z^2}$ and terms of order $\frac{J_\perp^2 h^2}{J_z \lambda^2}$, where the latter is what we want. So we further require $\frac{h^2}{\lambda^2} \gg \frac{J_\perp}{J_z}$ such that the term obtained

from the 3rd-order perturbation can be neglected.

Then we achieve the decorated BFG model

$$H_{eff} = -\frac{10J_\perp^2 h^2}{9J_z\lambda^2}\sum_{\bowtie}(\big| \quad \big\rangle\big\langle \quad \big| + h.c.) \tag{2.42}$$



in the parameter regime where $J_z \gg \lambda \gg J_\perp, h$ and $\frac{h^2}{\lambda^2} \gg \frac{J_\perp}{J_z}$.

2.7.2 Theorem-I as a special case of Theorem-II

Necessary condition for the existence of SRE state: constraints on the on-site projective representation First we prove that only when the on-site projective representation α satisfies the following 2 conditions is a SRE ground state possible.

1. $\alpha^N \simeq \mathbf{1} \in H^2(G, U(1))$.

2. $\gamma_g^\alpha(a) \simeq \mathbf{1} \in H^1(G, U(1))$.

Suppose the unit-cell is enlarged along x-direction to include N original unit-cell, then we have $T_x^N T_y T_x^{-N} T_y^{-1} = g^N = \mathbf{1}$,*i.e.*, we have usual translation T_x^N, T_y in the enlarged unit-cell. From Hastings' theorem we know that for a SRE ground state to exist, the enlarged unit-cell must carry usual representation. Hence we know α^N is a trivial 2-cocycle.

Next, we know from Theorem-II that for such a SRE state to exist, there must exist a 3-cocycle $\omega \in H^3(G, U(1))$, such that $\delta_g^\omega(a,b) = \alpha(a,b)^{-1}$ up to a 2-coboundary. By tuning the 2-coboundary of $\alpha(a,b)$, we are tuning the 1-coboundary of $\gamma_g^\alpha(a)$. Therefore we have $\gamma_g^\alpha(a) \simeq \gamma_g^\omega(a)$, where $\gamma_g^\omega(a) \equiv \frac{\delta_g^\omega(a,g)}{\delta_g^\omega(g,a)}$.

$$\begin{aligned}
\delta_g^\omega(a,g) &= \frac{\omega(a,g,g)\omega(g,a,g)}{\omega(a,g,g)} = \omega(g,a,g), \\
\delta_g^\omega(g,a) &= \frac{\omega(g,a,g)\omega(g,g,a)}{\omega(g,g,a)} = \omega(g,a,g).
\end{aligned} \tag{2.43}$$

Therefore we always have $\gamma_g^\omega(a) = 1$, which means $\gamma_g^\alpha(a) \simeq \mathbf{1} \in H^1(G, U(1))$.

Sufficient condition for the existence of SRE state: explicit construction of the 3-cocycle We shall show that the necessary condition given in the last section is also sufficient. To be more specific, we will construct a 3-cocycle $\omega \in H^3(G, U(1))$ out of α given $\alpha^N \simeq \mathbf{1} \in H^2(G, U(1))$ and $\gamma_g^\alpha(a) \simeq \mathbf{1} \in H^1(G, U(1))$, such that $\delta_g^\omega(a, b) = \alpha(a, b)^{-1}$. From Theorem-II, we know that such an SRE state described by 3-cocycle $\omega \in H^3(G, U(1))$ always exist, which completes our proof of Theorem-I.

Let's first fix a canonical gauge of $\alpha(a, b)$. Due to the direct product structure $G = G_1 \times Z_N$, we denote a general group element $a \in G$ as

$$a = g^{n_a} h_a, n_a = 0, 1 \cdots N-1, h_a \in G_1. \tag{2.44}$$

We are given the condition that α^N is a trivial 2-cocycle in $H^2(G, U(1))$. Let's first tune the 2-coboundary of $\alpha(a, b)$ such that $\alpha^N = 1$. Then $\alpha(a, b) \in Z_N$.

We also know that

$$\gamma_g^\alpha(a) \in B^1(G, U(1)). \tag{2.45}$$

If G_1 is a unitary group, then $\gamma_g^\alpha(a) \equiv 1$. If G_1 has anti-unitary operations, we should generally represent $\gamma_g^\alpha(a)$ as the 1-coboundary $\frac{\gamma}{{}^a\gamma}$.

Therefore we know that

$$\frac{\alpha(g, a)}{\alpha(a, g)} = \frac{\gamma}{{}^a\gamma} \in Z_N, \to \gamma \in Z_{2N}. \tag{2.46}$$

We choose the 2-coboundary $\delta(a) = \delta(g)^{n_a}$ where $\delta(g) = \gamma^{N-1}$.

Then under the 2-coboundary $\delta(a)$ we have

$$\alpha(a, b) \to \frac{\delta(g)^{n_a} \cdot^a (\delta(g)^{n_b})}{\delta(g)^{\langle n_a + n_b \rangle_N}} \alpha(a, b), \tag{2.47}$$

where $\langle n\rangle_N = n$ for $n < N$ and $\langle n\rangle_N = n - N$ for $n \geq N$.

Here the change of $\alpha(a,b)$ is always a Z_N element since

$$\begin{aligned}&\frac{\delta(g)^{n_a} \cdot^a (\delta(g)^{n_b})}{\delta(g)^{\langle n_a+n_b\rangle_N}}\\ &= \begin{cases} (\frac{{}^a\gamma}{\gamma})^{(N-1)n_b} \in Z_N, \text{if } n_a + n_b < N. \\ (\frac{{}^a\gamma}{\gamma})^{(N-1)n_b} \cdot \gamma^{N(N-1)} \in Z_N, \text{if } n_a + n_b \geq N. \end{cases}\end{aligned} \tag{2.48}$$

Then after the change of 2-coboundary, we still have $\alpha^N = 1$.

But $\gamma_g^\alpha(a)$ is changed as follows

$$\gamma_g^\alpha(a) = \frac{\alpha(g,a)}{\alpha(a,g)} \to \frac{\delta(g)}{{}^a\delta(g)} \cdot \frac{\gamma}{{}^a\gamma} = \frac{\gamma^N}{{}^a\gamma^N} = 1, \tag{2.49}$$

where we have used Eq. (2.48) and the fact that $\gamma^{2N} = 1$. Then after the change of 2-coboundary we always have $\gamma_g^\alpha(a) = 1$.

In summary, we have fixed $\alpha^N = 1$ and $\alpha(g,a) = \alpha(a,g)$ as the canonical gauge choice.

With the condition $\alpha^N = 1$ and $\alpha(g,a) = \alpha(a,g)$, we can explicitly construct the 3-cocycle as follows:

$$\omega(a,b,c) = [\alpha(b,c)^{-1}]^{n_a s_a}, s_a = 1/-1 \text{ for } a \text{ unitary/anti-unitary}. \tag{2.50}$$

First let's prove $\omega(a,b,c)$ is indeed a 3-cocycle. We have

$$\begin{aligned}&\omega(a,b,c)\omega(a,bc,d)\omega(b,c,d)^{s_a}\\ &= [\alpha(b,c)^{-1}]^{n_a s_a}[\alpha(bc,d)^{-1}]^{n_a s_a}[\alpha(c,d)^{-1}]^{n_b s_a s_b}\\ &= [\alpha(c,d)^{-1}]^{n_a s_a s_b}[\alpha(b,cd)^{-1}]^{n_a s_a}[\alpha(c,d)^{-1}]^{n_b s_a s_b},\end{aligned} \tag{2.51}$$

where in the last equality we have used the 2-cocycle condition of α, *i.e.*,

$$\alpha(b,c)\alpha(bc,d) = \alpha(c,d)^{s_b}\alpha(b,cd). \tag{2.52}$$

And we also have

$$\begin{aligned}&\omega(ab,c,d)\omega(a,b,cd)\\&=[\alpha(c,d)^{-1}]^{(\langle n_a+n_b\rangle_N)s_as_b}\cdot[\alpha(b,cd)^{-1}]^{n_as_a},\end{aligned}\tag{2.53}$$

which equals Eq. (2.51) since $\alpha^N=1$. Therefore ω satisfies the 3-cocycle condition

$$\omega(a,b,c)\omega(a,bc,d)\omega(b,c,d)^{s_a}=\omega(ab,c,d)\omega(a,b,cd).\tag{2.54}$$

Next we show that the slant product of ω with respect to g gives us α^{-1},

$$\begin{aligned}\delta_g^\omega(a,b)&=\frac{\omega(a,b,g)\omega(g,a,b)}{\omega(a,g,b)}=\frac{[\alpha(b,g)^{-1}]^{n_as_a}[\alpha(a,b)^{-1}]}{[\alpha(g,b)^{-1}]^{n_as_a}}\\&=\alpha(a,b)^{-1},\end{aligned}\tag{2.55}$$

where we have used $\alpha(b,g)=\alpha(g,b)$.

2.7.3 A brief introduction to symmetric tensor network representation of SPT phases

In this appendix we want to briefly summarize the symmetric tensor network representation of SPT phases and fix the notations for future convenience. More details of the general formalism can be found in Ref. [21, 47]. **Basic set-up** Let's consider a PEPS state formed by infinite numbers of site tensors and discuss the symmetry implementation on such state [40–47]. We assume that for a symmetric PEPS the symmetry transformed tensors and the original tensors are related by a gauge transformation:

$$W_g g\circ\mathbb{T}=\mathbb{T},\tag{2.56}$$

where $\mathbb{T}$ is the tensor states with all internal legs uncontracted and W_g is the product of the gauge transformation acting on all internal legs of the tensor network.

The invariant gauge group (IGG) is the group of all the gauge transformations leaving the uncontracted tensor $\mathbb{T}$ invariant. These are denoted as global IGG in contrast to the plaquette IGG introduced later. The global IGG naturally arises from the following tensor equations:

$$\mathbb{T} = W_a a W_b b \circ \mathbb{T} = W_{ab} ab \circ \mathbb{T}, \tag{2.57}$$

from which we know that

$$W_a \cdot^a W_b = \eta(a,b) W_{ab}, \tag{2.58}$$

where $\eta(a,b)$ should leave the tensor invariant and hence is an IGG element.

And we have the associativity condition for $\eta(a,b)$:

$$\eta(a,b)\eta(ab,c) =^{W_a a} \eta(b,c)\eta(a,bc). \tag{2.59}$$

The global IGG elements are a characteristic of symmetry breaking or topological order. In order to obtain an SPT state, we require all the global IGG elements can be decomposed into the product of plaquette IGG elements as shown in Fig 2.7, *i.e.*, $\eta(a,b) = \prod_p \lambda_p(a,b)$. There is a global phase ambiguity in such decomposition, namely we have $\prod_p \lambda_p = \prod_p \chi_p \lambda_p$ with χ_p a global phase since $\prod_p \chi_p = I$. The decomposible global IGG tells us that topological order is killed and the resulting phases should be an SPT phase described by the 3-cocycle ω, which arises as the phase ambiguity when lift Eq. (2.59) to plaquette IGG,

$$\lambda_p(a,b)\lambda_p(ab,c) = \omega_p(a,b,c)^{W_a a} \lambda_p(b,c)\lambda_p(a,bc). \tag{2.60}$$

The ω shown above is a well-defined 3-cocycle since the phase ambiguities in λ will only modify it by a 3-coboundary.

Representation of $\delta_g^\omega(a,b)$ using plaquette IGG In this subsection we want to give a representation of the slant product $\delta_g^\omega(a,b)$ in terms of plaquette IGG for an SPT state

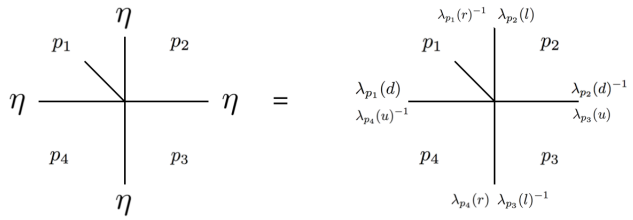


Figure 2.7: The decomposition of global IGG into plaquette IGG. λ's from different plaquettes commutes with each other, and the action of any two λ's in the same plaquette leave the tensor invariant.

characterized by 3-cocycle ω, where g lies in the center of the whole symmetry group G. First, from definition we have

$$\delta_g^\omega(a,b) = \frac{\omega(a,b,g)\omega(g,a,b)}{\omega(a,g,b)} \tag{2.61}$$

The 3-cocycle arises from the decomposition of global IGG into the plaquette IGG, see Eq. (2.60). Therefore in order to compute $\delta_g^\omega(a,b)$, we need the following equations:

$$\begin{aligned}
\lambda_p(a,b)\cdot\lambda_p(ab,g) &= \omega(a,b,g)\cdot^a \lambda_p(b,g)\cdot\lambda_p(a,bg),\\
\lambda_p(g,a)\cdot\lambda_p(ga,b) &= \omega(g,a,b)\cdot^g \lambda_p(a,b)\cdot\lambda_p(g,ab),\\
\lambda_p(a,g)\cdot\lambda_p(ag,b) &= \omega(a,g,b)\cdot^a \lambda_p(g,b)\cdot\lambda_p(a,gb).
\end{aligned} \tag{2.62}$$

Writing Eq. (2.62) in a more convenient way (we ignore the subscript p henceforth):

$$\begin{aligned}
\omega(a,b,g) &= \lambda^{-1}(a,bg)\cdot^{W_a a}\lambda^{-1}(b,g)\cdot\lambda(a,b)\cdot\lambda(ab,g),\\
\omega(g,a,b) &= \lambda^{-1}(g,ab)\cdot^{W_g g}\lambda^{-1}(a,b)\cdot\lambda(g,a)\cdot\lambda(ga,b),\\
\omega^{-1}(a,g,b) &= \lambda^{-1}(ag,b)\lambda^{-1}(a,g)\cdot^{W_a a}\lambda(g,b)\cdot\lambda(a,gb).
\end{aligned} \tag{2.63}$$

We have

$$\begin{aligned}\delta_g^\omega(a,b) &= [\omega(g,a,b)]\cdot[\omega^{-1}(a,g,b)]\cdot[\omega(a,b,g)]\\ &= \lambda^{-1}(g,ab)\cdot^{W_{gg}}\lambda^{-1}(a,b)\cdot[\lambda(g,a)\cdot\lambda^{-1}(a,g)]\\ &\cdot^{W_{aa}}[\lambda(g,b)\cdot\lambda^{-1}(b,g)]\cdot\lambda(a,b)\cdot\lambda(ab,g)\end{aligned} \tag{2.64}$$

We can simplify Eq. (2.64) by defining ${}^{W_{gg}}W_a a = \xi_a(g)W_a a, a \in G$, where $\xi_a(g) = \prod\lambda_a(g)$. Another way of computing $\xi_a(g)$ is

$$\begin{aligned}&\xi_a(g) = W_g g W_a a (W_g g)^{-1}(W_a a)^{-1} = \eta(g,a)\eta^{-1}(a,g)\\ &\rightarrow \lambda_a(g) = \lambda(g,a)\lambda^{-1}(a,g).\end{aligned} \tag{2.65}$$

Then Eq. (2.64) becomes

$$\lambda_a(g)\cdot^{W_{aa}}\lambda_b(g) = \delta_g^\omega(a,b)\cdot^{W_{gg}}\lambda(a,b)\cdot\lambda_{ab}(g)\cdot\lambda^{-1}(a,b). \tag{2.66}$$

2.7.4 The projective representation carried by a g-symmetry-defect

In this section we want to give a tensor proof of the following fact [48]: for an SPT state characterized by the 3-cocycle $\omega(a,b,c) \in H^3(G,U(1))$, the projective representation carried by the symmetry g-defect is represented by the inverse of the slant product $[\delta_g^\omega(a,b)]^{-1}$.

To this end, we first create an open g-defect string with a pair of g-defects on the two ends in the given ground-state SPT wave-function $|\Psi\rangle$. The wave-function is denoted as $|\Psi_{defect}\rangle$. This is done in the tensor language by inserting W_g strings and modifying the tensors close to the defect core as shown in Fig. 2.8.

Let's take a patch enclosing one of the two g-defects and measure the projective representation carried by the g-defect. Before the insertion of the g-defect, the local symmetry

action $U(a)$ on the patch is defined as acting W_a on the virtual legs on the edge and $D(a)$ on the physical legs inside the patch, *i.e.*,

$$U(a) = \prod_{\text{boundary}} W_a \prod_{\text{bulk}} D(a). \tag{2.67}$$

The symmetry operation should leave $|\Psi\rangle$ invariant up to a phase. Then the projective representation inside the patch is measured by acting

$$D(a) \cdot D(b) \cdot [D(a \circ b)]^{-1} \tag{2.68}$$

on the physical legs inside the patch. Alternatively, we can do this by monitoring the inverse of the phase generated by acting $\eta(a,b) \equiv W_a \cdot^a W_b \cdot (W_{ab})^{-1}$ on the boundary virtual legs since, by our assumption, the action of $U(a)U(b)U(ab)^{-1}$ leaves the patch fully invariant.

In general , acting $\eta(a,b)$ on the virtual legs of a tensor leaves the tensor invariant only up to a phase. Therefore η itself is not a global IGG. Instead, we have

$$\eta(a,b) = W_x(a,b)\eta'(a,b), \tag{2.69}$$

where $W_x(a,b)$ is a pure-phase gauge transformation which yields the extra phase for each site and $\eta'(a,b)$ leaves every tensor invariant. Now $\eta'(a,b)$ is decomposable and we denote it as $\eta'(a,b) = \prod \lambda_p(a,b)$.

As for our present case, suppose we have the action of $\prod_{\text{boundary}} \eta(a,b)$ on the ground state wave-function

$$\prod_{\text{boundary}} \eta(a,b) \, |\Psi\rangle = e^{i\phi} \, |\Psi\rangle \,, \tag{2.70}$$

from which we know that the projective representation inside the patch is just $\prod_{\text{bulk}}(D(a) \cdot D(b) \cdot [D(a \circ b)]^{-1}) = e^{-i\phi}$.

From the previous discussion we have

$$e^{-i\phi}\prod_{\text{boundary}}\eta(a,b)=\prod_{\text{boundary}}\eta'(a,b). \tag{2.71}$$

After the insertion of g-defect, the new local symmetry operation $U^g(a)$ should be defined as acting $\tilde{W}_a$ on the virtual legs on the edge and $D(a)$ on the physical legs inside the patch. Here we have used the $\tilde{W}_a$ as shown in Fig. 2.9 where $\tilde{W}_a$ on the bond crossing the defect line is changed to be $[\lambda_a(g)](d)\cdot W_a$ and remains W_a elsewhere. This newly-defined $\tilde{W}_a$ ensures that no boundary excitations are created. The physical symmetry operation in the bulk should still be the same.

We will then use $\prod_{\text{boundary}}\tilde{\eta}(a,b)$ to measure the projective representation inside the patch, where $\tilde{\eta}(a,b)\equiv\tilde{W}_a\cdot^a\tilde{W}_b\cdot(\tilde{W}_{ab})^{-1}$. Far away from the defect core, the tensor wave-function is basically the same as before. And $\prod_{\text{boundary}}\tilde{\eta}(a,b)$ will be the same as $\prod_{\text{boundary}}\eta(a,b)$ except at the bond crossing the defect line, see Fig. 2.10. The $\tilde{\eta}(a,b)$ at the bond crossing the defect line should be $\tilde{W}_a\cdot^a\tilde{W}_b\cdot(W_{ab})^{-1}$:

$$\begin{aligned}
&[\lambda_a(g)](d)\cdot W_a\cdot^a[\lambda_b(g)](d)\cdot^a W_b\cdot([\lambda_{ab}(g)](d)\cdot W_{ab})^{-1}\\
&=[\lambda_a(g)](d)\cdot^{W_a a}[\lambda_b(g)](d)\cdot W_a\cdot^a W_b\cdot(W_{ab})^{-1}[\lambda_{ab}^{-1}(g)](d)\\
&=[\lambda_a(g)](d)\cdot^{W_a a}[\lambda_b(g)](d)\cdot\eta(a,b)\cdot[\lambda_{ab}^{-1}(g)](d).
\end{aligned} \tag{2.72}$$

We can work with the decomposable $\eta'(a,b)$ at the boundary if we keep track of the $e^{i\phi}$ phase. After going through the calculation as shown in Fig. 2.10, we have

$$\prod_{bdr}\tilde{\eta}(a,b)\,|\Psi_{defect}\rangle=e^{i\phi}\delta_g^{\omega}(a,b)\,|\Psi_{defect}\rangle\,. \tag{2.73}$$

Comparing Eq. (2.70) with Eq. (2.73) and note that projective representation is defined through the inverse of Eq. (2.70), we know that the g-defect carries $[\delta_g^{\omega}(a,b)]^{-1}$ projective representation.

2.7.5 Consequence of the magnetic translation symmetry in tensor-network formulation

The magnetic translation symmetry and on-site projective representation constrain the possible symmetric short-range entangled states in a system. Specifically, we have the following fact: for a system with on-site projective representation of the on-site symmetry group G characterized by $\alpha(a,b)$ and magnetic translation symmetry $T_xT_yT_x^{-1}T_y^{-1}=g$, an SPT ground state described by the 3-cocycle $\omega\in H^3(G,U(1))$ can be realized as its ground state only when $\delta_g^{\omega}(a,b)=\alpha(a,b)^{-1}$.

Basic set-up We have $T_xT_yT_x^{-1}T_y^{-1}=g$, which leads to

$$W_{T_x}T_xW_{T_y}T_y(W_{T_x}T_x)^{-1}(W_{T_y}T_y)^{-1}=W_gg. \tag{2.74}$$

We define $\eta(a,b)$ as

$$W_a\cdot^a W_b=\eta(a,b)W_{ab}. \tag{2.75}$$

By acting $\eta(a,b)$ on $\mathbb{T}$, we will get an extra phase $\alpha(a,b)^{-1}$ per unit-cell, therefore it is not a global *IGG*. However, we can define a pure-phase gauge transformation $W(\alpha(a,b))$ which yields exactly the phase $\alpha(a,b)^{-1}$ for every site tensor, see Fig. for an illustration.

Then we can write $\eta(a,b)$ as

$$\eta(a,b)=W(\alpha(a,b))\eta'(a,b), \tag{2.76}$$

where $\eta'(a,b)=\prod\lambda(a,b)$ is an IGG and is decomposable.

Similarly since we have $D(g)\circ D(a)=\gamma_a(g)D(a)\circ D(g)$ on physical leg, where $\gamma_a(g)=\alpha(g,a)/\alpha(a,g)\in H^1(G,U(1))$. We write the action of W_gg on W_aa as

$${}^{W_gg}W_aa=W(\gamma_a(g))\xi_a(g)\cdot W_aa, \tag{2.77}$$

where $\xi_g(a) = \prod \lambda_g(a)$ is a decomposable IGG element and the definition of $W(\gamma_a(g))$ is the same as $W(\alpha(a,b))$.

And as for T_x, T_y, we have

$$^{W_{T_x}T_x}W_a a = W(\gamma_a(T_x))\xi_a(T_x)W_a a, etc.. \tag{2.78}$$

We act Eq. (2.74) on $W_a a$ and obtain an IGG equation

$$^{W_g g}\xi_a^{-1}(T_y) \cdot^{W_g g W_{T_y}T_y} \xi_a^{-1}(T_x) \cdot^{W_{T_x}T_x} \xi_a(T_y) \cdot \xi_a(T_x) = \xi_a(g), \tag{2.79}$$

which, when lift to plaquette IGG, should give us (we have absorbed the phase ambiguity into the definition of $\lambda_a(g)$)

$$^{W_g g}\lambda_a^{-1}(T_y) \cdot^{W_g g W_{T_y}T_y} \lambda_a^{-1}(T_x) \cdot^{W_{T_x}T_x} \lambda_a(T_y) \cdot \lambda_a(T_x) = \lambda_a(g). \tag{2.80}$$

Acting $W_g g$ We first act $W_g g$ on Eq. (2.75), then we have

$$\begin{aligned} &^{W_g g}[W_a a W_b b] =^{W_g g} [\eta(a,b)W_{ab}ab] \\ &\Rightarrow W(\gamma_a(g)) \cdot^{a} W(\gamma_b(g))\xi_a(g) \cdot^{W_a a} \xi_b(g) \cdot W_a a W_b b \\ &=^{W_g g} \eta(a,b) \cdot W(\gamma_{ab}(g))\xi_{ab}(g)W_{ab}ab, \end{aligned} \tag{2.81}$$

which then leads to

$$\xi_a(g) \cdot^{W_a a} \xi_b(g) =^{W_g g} \eta'(a,b) \cdot \xi_{ab}(g) \cdot \eta'^{-1}(a,b), \tag{2.82}$$

where extra phase factors $W(\alpha(a,b))$, $W(\gamma_a(g))$ all cancel.

When lifting Eq. (2.82) to plaquette, we have (from Eq. (2.66))

$$\lambda_a(g) \cdot^{W_a a} \lambda_b(g) = \delta_g^{\omega}(a,b) \cdot^{W_g g} \lambda(a,b) \cdot \lambda_{ab}(g) \cdot \lambda(a,b)^{-1}. \tag{2.83}$$

Acting translation We have another way of deriving Eq. (2.83). We first act $W_{T_x}T_x$ on the two sides of Eq. (2.75) and obtain an IGG equation,

$$\xi_a(T_x) \cdot^{W_a a} \xi_b(T_x) =^{W_{T_x}T_x} \eta(a,b) \cdot \xi_{ab}(T_x)\eta(a,b)^{-1}, \tag{2.84}$$

where the extra $W(\gamma_a(T_x)), W(\gamma_b(T_x)), W(\gamma_{ab}(T_x))$ cancel since we have $\gamma_a(T_x) \cdot^a \gamma_b(T_x) = \gamma_{ab}(T_x)$.

When lift Eq. (2.84) to plaquette IGG, we have

$$\lambda_a(T_x) \cdot^{W_a a} \lambda_b(T_x) = [\alpha(a,b)]^{-y} \cdot^{W_{T_x} T_x} \lambda(a,b)\lambda_{ab}(T_x)\lambda^{-1}(a,b), \tag{2.85}$$

where $\prod[\alpha(a,b)]^{-y} =^{W_{T_x} T_x} W(\alpha(a,b)) \cdot W(\alpha(a,b))^{-1}$ and the plaquette IGG $[\alpha(a,b)]^{-y}$ are just loops of phases as shown in Fig. 2.12.

Similarly we have

$$\lambda_a(T_y) \cdot^{W_a a} \lambda_b(T_y) =^{W_{T_y} T_y} \lambda(a,b)\lambda_{ab}(T_y)\lambda^{-1}(a,b), \tag{2.86}$$

where there is no extra factor coming from ${}^{W_{T_y} T_y} W(\alpha(a,b)) \cdot W(\alpha(a,b))^{-1}$ since it is T_y invariant.

With Eq. (2.85) and Eq. (2.86) we can explicitly calculate the action of

$$W_{T_x} T_x W_{T_y} T_y (W_{T_x} T_x)^{-1} (W_{T_y} T_y)^{-1}$$

on Eq. (2.75) in terms of plaquette IGG. The action of $W_{T_x} T_x W_{T_y} T_y (W_{T_x} T_x)^{-1} (W_{T_y} T_y)^{-1}$

on LHS of Eq. (2.75) is

$$
\begin{aligned}
&{}^{W_g g}[\lambda_a(T_y) \cdot^{W_a a} \lambda_b(T_y)]^{-1} \cdot^{W_g g W_{T_y} T_y} [\lambda_a(T_x)^{W_a a} \lambda_b(T_x)]^{-1} \\
&\cdot^{W_{T_x} T_x} [\lambda_a(T_y)^{W_a a} \lambda_b(T_y)] \cdot [\lambda_a(T_x)^{W_a a} \lambda_b(T_x)] \\
&=^{W_g g W_a a} \lambda_b^{-1}(T_y) \cdot^{W_g g} [\lambda_a^{-1}(T_y) \cdot^{W_{T_y} T_y W_a a} \lambda_b^{-1}(T_x)^{W_{T_y} T_y} \lambda_a^{-1}(T_x)] \\
&\cdot^{W_{T_x} T_x} \lambda_a(T_y) \cdot^{W_{T_x} T_x W_a a} \lambda_b(T_y) \cdot \lambda_a(T_x)^{W_a a} \lambda_b(T_x) \\
&= \lambda_a(g)^{W_a a W_g g} \lambda_b^{-1}(T_y) \cdot \lambda_a^{-1}(g) \cdot^{W_g g} [^{W_a a W_{T_y} T_y} \lambda_b^{-1}(T_x) \lambda_a^{-1}(T_y)^{W_{T_y} T_y} \lambda_a^{-1}(T_x)] \\
&\cdot^{W_{T_x} T_x} \lambda_a(T_y) \cdot \lambda_a(T_x) \\
&\cdot^{W_a a W_{T_x} T_x} \lambda_b(T_y) \cdot^{W_a a} \lambda_b(T_x) \\
&= \lambda_a(g) \cdot^{W_a a W_g g} \lambda_b^{-1}(T_y) \cdot^{W_a a W_g g W_{T_y} T_y} \lambda_b^{-1}(T_x) \cdot [\lambda_a^{-1}(g)^{W_g g} \lambda_a^{-1}(T_y)^{W_g g W_{T_y} T_y} \lambda_a^{-1}(T_x) \\
&\cdot^{W_{T_x} T_x} \lambda_a(T_y) \lambda_a(T_x)] \\
&\cdot^{W_a a W_{T_x} T_x} \lambda_b(T_y)^{W_a a} \lambda_b(T_x) \\
&= \lambda_a(g) \cdot^{W_a a} [^{W_g g} \lambda_b^{-1}(T_y) \cdot^{W_g g W_{T_y} T_y} \lambda_b^{-1}(T_x) \cdot^{W_{T_x} T_x} \lambda_b(T_y) \lambda_b(T_x)] \\
&= \lambda_a(g) \cdot^{W_a a} \lambda_b(g),
\end{aligned}
\tag{2.87}
$$

where we have used Eq. (2.77) and Eq. (2.78) in the first three equalities and we have used Eq. (2.80) in the last two equalities.

The action of $W_{T_x}T_xW_{T_y}T_y(W_{T_x}T_x)^{-1}(W_{T_y}T_y)^{-1}$ on RHS of Eq. (2.75) is

$$
\begin{aligned}
&{}^{W_gg}[{}^{W_{T_y}T_y}\lambda(a,b)\lambda_{ab}(T_y)\lambda^{-1}(a,b)]^{-1}\cdot{}^{W_ggW_{T_y}T_y}[\alpha(a,b)^{-y}\\
&\cdot{}^{W_{T_x}T_x}\lambda(a,b)\lambda_{ab}(T_x)\lambda^{-1}(a,b)]^{-1}\\
&\cdot{}^{W_{T_x}T_x}[{}^{W_{T_y}T_y}\lambda(a,b)\lambda_{ab}(T_y)\lambda^{-1}(a,b)]\cdot\alpha(a,b)^{-y}\cdot{}^{W_{T_x}T_x}\lambda(a,b)\lambda_{ab}(T_x)\lambda^{-1}(a,b)\\
&={}^{W_{T_y}T_y}[\alpha(a,b)^{-y}]^{-1}\cdot\alpha(a,b)^{-y}\cdot{}^{W_gg}\lambda(a,b)^{W_gg}\lambda_{ab}^{-1}(T_y)\\
&\cdot{}^{W_ggW_{T_y}T_y}\lambda_{ab}^{-1}(T_x)\cdot{}^{W_{T_x}T_x}\lambda_{ab}(T_y)\cdot\lambda_{ab}(T_x)\cdot\lambda^{-1}(a,b)\\
&=\alpha(a,b)^{-1}\cdot{}^{W_gg}\lambda(a,b)\cdot\lambda_{ab}(g)\cdot\lambda^{-1}(a,b),
\end{aligned}
\tag{2.88}
$$

where we have used Eq. (2.80) and $\alpha^{-1}(a,b)$ is just a plaquette IGG with loop of phases $\alpha^{-1}(a,b)$.

Combining Eq. (2.87) with Eq. (2.88), we have

$$\lambda_a(g)\cdot{}^{W_{a}a}\lambda_b(g)=\alpha^{-1}(a,b)^{W_gg}\lambda(a,b)\cdot\lambda_{ab}(g)\cdot\lambda^{-1}(a,b). \tag{2.89}$$

Comparing Eq. (2.89) with Eq. (2.83), we have

$$\alpha^{-1}(a,b)=\delta_g^{\omega}(a,b). \tag{2.90}$$

It is easy to see that the global phase ambiguities in Eq. (2.80),(2.85),(2.86) will at most modify the LHS of Eq. (2.90) up to a 2-cobounday, therefore it should be understood as the 2-cycle equivalence $\delta_g^{\omega}\simeq\alpha^{-1}\in H^2(G,U(1))$.

2.7.6 Generic constructions of symmetry-enforced SPT tensor-network wavefunctions

In this section we want to construct an SPT state with on-site symmetry group G, magnetic translation symmetry satisfying $T_xT_yT_x^{-1}T_y^{-1}=g$ where the SPT is characterized by the

3-cocycle $\omega \in H^3(G, U(1))$ and the on-site symmetry group is represented projectively with the 2-cocycle $\alpha(a, b)$ equal to the inverse of the slant product of ω with respect to g,*i.e.*,

$$\alpha(a, b) \simeq [\delta_g^\omega(a, b)]^{-1} \in H^2(G, U(1)). \tag{2.91}$$

To achieve this goal, we will use the tensor network formalism. Let's start from a SPT tensor wavefunction with the symmetry group $\mathbb{Z}^2 \times G$ described by a 3-cocycle $\omega \in H^3(G, U(1))$, where $\mathbb{Z}^2$ represents the usual translation $T_x^{orig.}, T_y^{orig.}$. Then we know that every tensor is invariant under the action $D(a)$ on physical leg together with $\prod W_a$ on all the virtual legs, from which we have a set of tensor equations. Here we require $D(a)$ to be a direct sum of usual representation $D_1(a)$ and projective representation $D_2(a)$ with 2-cocycle $[\delta_g^\omega(a, b)]^{-1}$. We choose the gauge transformation $W_{T_x^{orig.}}, W_{T_y^{orig.}}$ to be identity for simplicity. The global IGG $\eta(a, b)$ comes from the following tensor equation:

$$W_a \cdot^a W_b = \eta(a, b) W_{ab}, \tag{2.92}$$

where $\eta(a, b)$ is decomposable, *i.e.*, $\eta(a, b) = \prod \lambda_p(a, b)$. We require tensors to be fully invariant under $\eta(a, b)$ without even generating phases. This condition ensures $D(a)$ on the physical legs to be projected onto $D_1(a)$ sector.

We define ${}^{W_g g}W_a a = \xi_a(g) W_a a$, where $\xi_a(g) = \prod \lambda_a(g)$ is a decomposable global IGG. More generally we define ${}^{(W_g g)^x}W_a = \xi_a(g, x) W_a$, where $\xi_a(g, x) = \prod \lambda_a(g, x)$. Then we have the following relation:

$$\lambda_a(g, x + 1) = {}^{W_g g}\, \lambda_a(g, x) \cdot \lambda_a(g). \tag{2.93}$$

Then we change our tensor wave-function in the following way:

1. We revise the original tensor such that it is invariant under the symmetry operation and the new plaquette IGG defined in Fig. 2.13. Note that this step is necessary for

us to obtain a symmetric and non-vanishing tensor wave-function after the insertion of W_g.

2. We insert $[W_g(u)]^x$ on the upper leg of every tensor as shown in Fig. 2.14. Physically it means inserting one g-defect per unit-cell.

3. We define the new on-site symmetry operation $\tilde{W}_a$ and translation symmetry W_{T_x} as shown in Fig. 2.14. We will show that the new tensor is invariant under such symmetry transformations.

From the last section we have shown that every g-defect carries a projective representation represented by $\delta_g^\omega(a,b)^{-1}$, therefore one would expect after insertion of g-defects, we now have one $\delta_g^\omega(a,b)^{-1}$ projective representation per unit-cell. Let's show it more clearly through explicit calculations.

First, let's show that the revised tensor wave-function satisfies all the required symmetries. It's apparent that the new tensor has magnetic translation symmetry $W_{T_x}T_x, W_{T_y}T_y$ defined in Fig. 2.14, *i.e.*

$$W_{T_x}T_xW_{T_y}T_y(W_{T_x}T_x)^{-1}(W_{T_y}T_y)^{-1} = W_g g. \tag{2.94}$$

It can be proven that the new tensor is invariant under the new symmetry transformation $\tilde{W}_a a$ as shown in Fig. 2.15. The invariance of the new tensor $\tilde{T}^{(x,y)}$ under the new plaquette IGG is also apparent as shown in Fig. 2.14. Then we know that Eq. (2.60) still holds for this state, which means the new state obtained is still an SPT state described by the same 3-cocycle ω.

Finally we want to show that the new tensor now carries a projective representation $[\delta_g^\omega(a,b)]^{-1}$ per unit-cell. It can be proven that by acting $\tilde{W}_a \cdot^a \tilde{W}_b \cdot \tilde{W}_{ab}^{-1}$ on all the virtual

legs of a tensor we will get a phase $\delta_g^\omega(a,b)$ for every tensor as shown in Fig. 2.16.

In doing so we need the following identity

$$\lambda_a(g,x)^{W_a a}\lambda_b(g,x)\lambda(a,b)\lambda_{ab}(g,x)^{-1} \cdot^{(W_g g)^x} [\lambda(a,b)^{-1}] = [\delta_g^\omega(a,b)]^x. \tag{2.95}$$

Let's denote the LHS of Eq. (2.95) as $f(x)$. From Eq. (2.66) we have $f(1) = \delta_g^\omega(a,b)$, then we need to find out the relation between $f(x)$ and $f(x+1)$. Using $\lambda(g,x) =^{W_g g} \lambda(g,x-1)\cdot\lambda(g)$, we can rewrite Eq. (2.95) as

$$\begin{aligned}
&[^{W_g g}\lambda_a(g,x-1)\lambda_a(g)]^{W_a a}[^{W_g g}\lambda_b(g,x-1)\cdot\lambda_b(g)]\lambda(a,b)[^{W_g g}\lambda_{ab}(g,x-1)\cdot\lambda_{ab}(g)]^{-1} \cdot^{(W_g g)^x} [\lambda(a,b)^{-1}] \\
&= [^{W_g g}\lambda_a(g,x-1)\lambda_a(g)]^{W_a a W_g g}\lambda_b(g,x-1) \cdot^{W_a a} \lambda_b(g)\lambda(a,b)\lambda_{ab}(g)^{-1} \\
&\cdot [^{W_g g}\lambda_{ab}(g,x-1)]^{-1} \cdot^{(W_g g)^x} [\lambda(a,b)^{-1}] \\
&=^{W_g g} \lambda_a(g,x-1)[\lambda_a(g)\xi_a^{-1}(g)]^{W_g g W_a a}\lambda_b(g,x-1)\xi_a(g)^{W_a a}\lambda_b(g)\lambda(a,b)\lambda_{ab}(g)^{-1} \\
&\cdot^{W_g g} [\lambda_{ab}(g,x-1)]^{-1} \cdot^{(W_g g)^x} [\lambda(a,b)^{-1}] \\
&= \delta_g^\omega(a,b)^{W_g g}\lambda_a(g,x-1)[\lambda_a(g)\xi_a^{-1}(g)]^{W_g g W_a a}\lambda_b(g,x-1)[\xi_a(g)\lambda_a(g)^{-1}]^{W_g g}\lambda(a,b) \\
&\cdot^{W_g g} [\lambda_{ab}(g,x-1)]^{-1} \cdot^{(W_g g)^x} [\lambda(a,b)^{-1}] \\
&= \delta_g^\omega(a,b)^{W_g g}[\lambda_a(g,x-1)^{W_a a}\lambda_b(g,x-1)\lambda(a,b)\cdot\lambda_{ab}(g,x-1)^{-1} \cdot^{(W_g g)^{x-1}} \lambda(a,b)^{-1}].
\end{aligned} \tag{2.96}$$

The above derivation tells us that $f(x) = \delta_g^\omega(a,b)^{W_g g} f(x-1)$, therefore by induction we have $f(x) = [\delta_g^\omega(a,b)]^x$.

With the help of Eq. (2.95), we can readily calculate the new IGG $\tilde{\eta}(a,b) \equiv \tilde{W}_a \cdot^a \tilde{W}_b \cdot \tilde{W}_{ab}^{-1}$. The $\tilde{\eta}(a,b)$ on the up and down virtual legs are just $\eta(a,b)$ defined before. The

$\tilde{\eta}(a,b)$ on the left leg is computed as follows:

$$\begin{aligned}
&[\lambda_a(g,-x)](d)\cdot W_a \cdot^a [\lambda_b(g,-x)](d) \cdot^a W_b \cdot ([\lambda_{ab}(g,-x)](d)\cdot W_{ab})^{-1} \\
&= [\lambda_a(g,-x)](d) \cdot^{W_a a} [\lambda_b(g,-x)](d)\cdot W_a \cdot^a W_b \cdot (W_{ab})^{-1}[\lambda_{ab}(g,-x)](d)^{-1} \\
&= [\lambda_a(g,-x)](d) \cdot^{W_a a} [\lambda_b(g,-x)](d)\cdot \eta(a,b)\cdot [\lambda_{ab}(g,-x)](d)^{-1} \\
&= [\delta_g^\omega(a,b)]^{-x}[\lambda(a,b)](u)^{-1} \cdot^{(W_g g)^{-x}} [\lambda(a,b)](d).
\end{aligned} \tag{2.97}$$

Similarly, we have on the right leg

$$[\delta_g^\omega(a,b)]^{x+1} \cdot^{(W_g g)^{-x-1}} [\lambda(a,b)](d)^{-1}\cdot[\lambda(a,b)](u). \tag{2.98}$$

Therefore, as shown in Fig. 2.16, we know that $\tilde{\eta}(a,b)$ is just $\delta_g^\omega(a,b)$ times the product of plaquette IGG shown in Fig. 2.14 which leaves the tensor invariant up to a phase $\delta_g^\omega(a,b)$. Then on the physical leg we are forced to have $D(a)D(b) = [\delta_g^\omega(a,b)]^{-1}D(ab)$, *i.e.*, $D(a)$ is projected onto $D_2(a)$ sector. So the desired on-site projective representation is achieved.

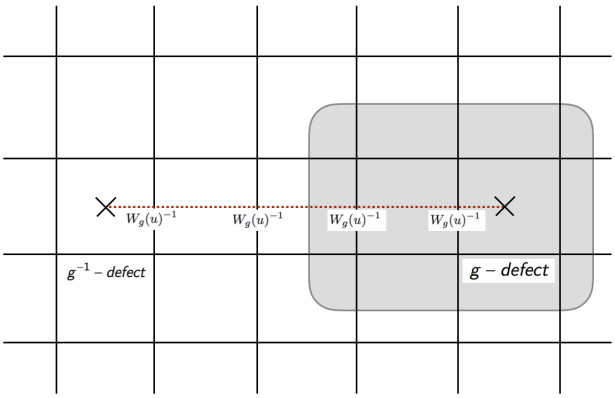


Figure 2.8: An example of g-defect line. The g-defect line is obtained by inserting W_g on only one side of the virtual legs crossed by the red dashed line. The tensors close to the defect core should be revised in order to make the tensor wave-function symmetric and non-vanishing. Following the usual convention, we say that the defect line always points from g^{-1}-defect to g-defect, and we always insert W_g to the left when one goes forward along the line. Therefore in the figure we can identify the right end as the g-defect (remember $W_g(d) = W_g(u)^{-1}$). The grey area encloses a g-defect and we can to measure its projective representation through the action of $\eta'(a, b)$ on the boundary virtual legs, see the discussion in the main text.

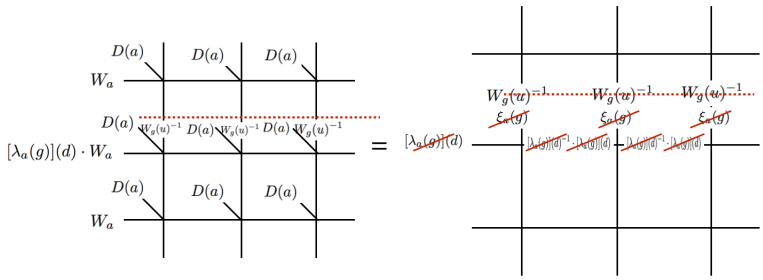


Figure 2.9: Invariance of the wave-function under $U^g(a)$. In the figure we can see that $\tilde{W}_a = [\lambda_a(g)](d) \cdot W_a$ where the defect line crosses the boundary and $\tilde{W}_a = W_a$ elsewhere. Such a definition ensures that no boundary excitations are created by acting $U^g(a)$ (for the moment we do not care about what happens at the defect core). In deriving the second figure, we have used the invariance of the tensor under $W_a a$, the identity $W_a W_g^{-1} = W_g^{-1} \xi_a(g) W_a$ and invariance of the tensor under plaquette IGG $\lambda_a(g)$.

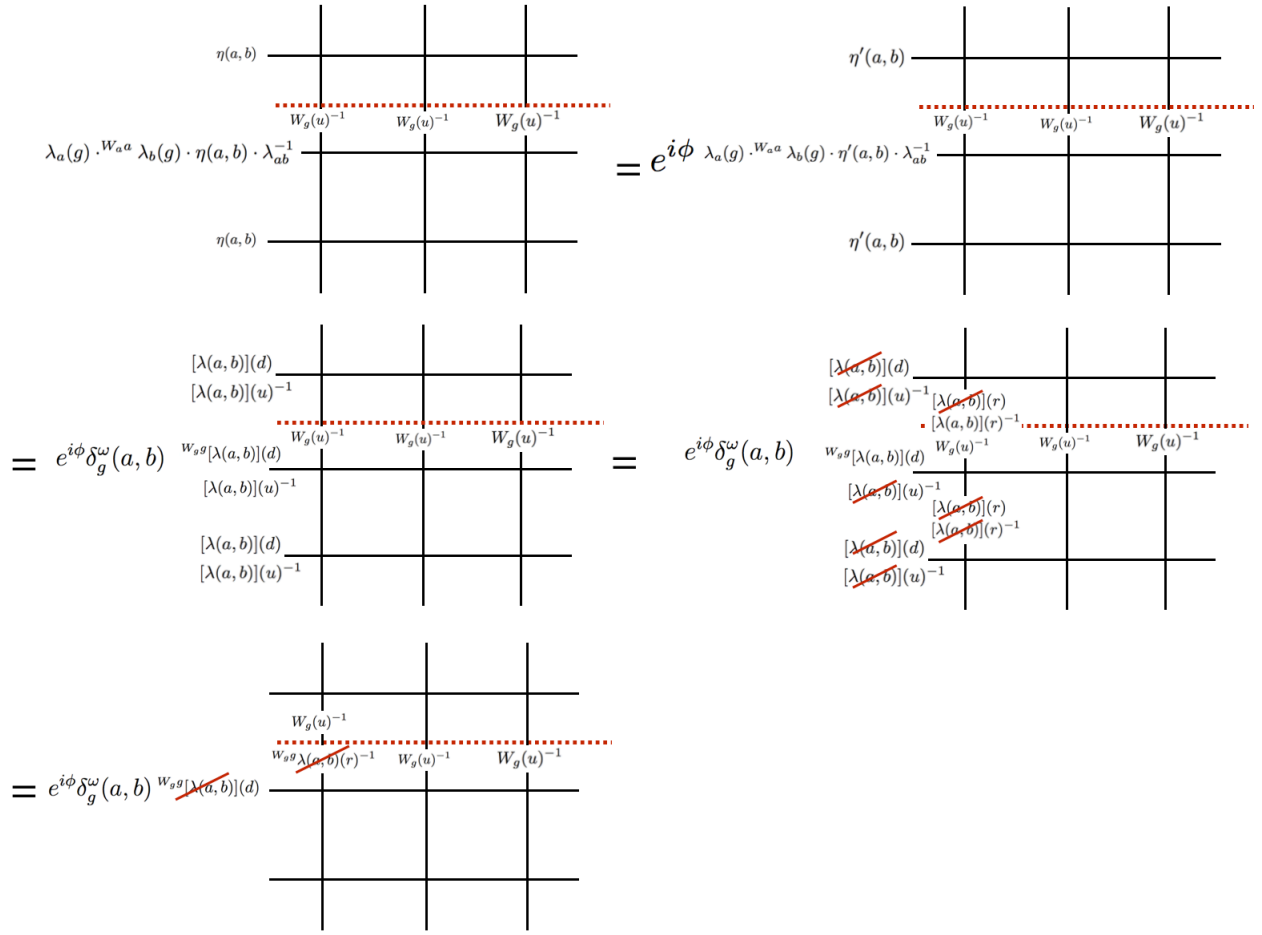


Figure 2.10: Measurement of projective representation carried by g-defect. From Eq. (2.72), we know that $\tilde{\eta}(a,b) = \lambda_a(g) \cdot^{W_a a} \lambda_b(g) \cdot \eta(a,b) \cdot \lambda_{ab}^{-1}$ where the boundary is crossed by the defect line and $\tilde{\eta}(a,b) = \eta(a,b)$ elsewhere. In the first equality we have used Eq. (2.71). In the second equality we have used the decomposition of $\eta'(a,b)$ and Eq. (2.66). In the third equality we have used the tensor invariance under plaquette IGG. In the last equality we have used the identity $\lambda(r)^{-1} \cdot W_g^{-1} = W_g^{-1} \cdot^{W_g} \lambda(r)^{-1}$ and the tensor invariance under plaquette IGG.

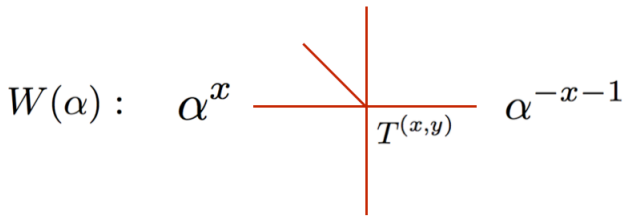


Figure 2.11: The definition of phase-gauge transformation $W(\alpha(a,b))$.

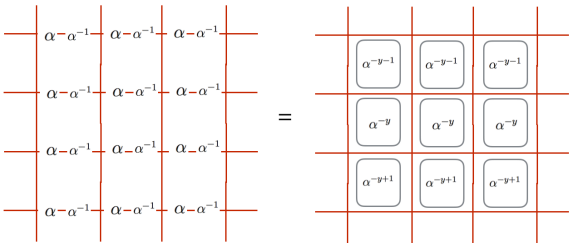


Figure 2.12: The decomposition rule of ${}^{W_{T_x}T_x}W(\alpha(a,b)) \cdot W(\alpha(a,b))^{-1}$ (LHS) as a product of plaquette IGG $\lambda(\alpha(a,b))$ (RHS).

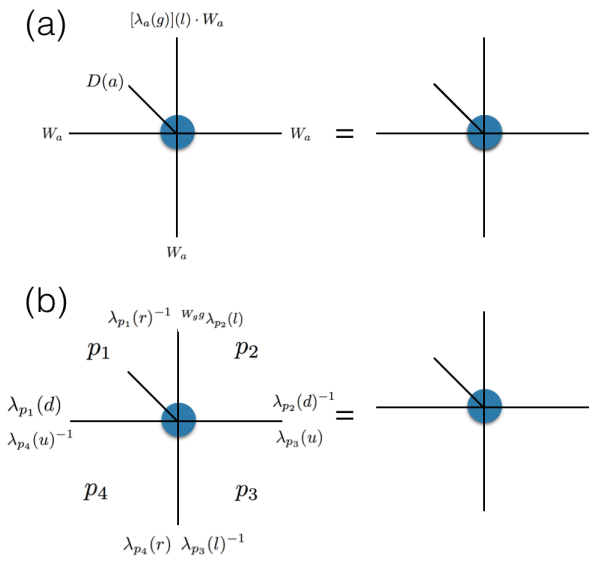


Figure 2.13: The original tensor before insertion of g-defect is required to be invariant under the revised symmetry operation and the revised plaquette IGG.

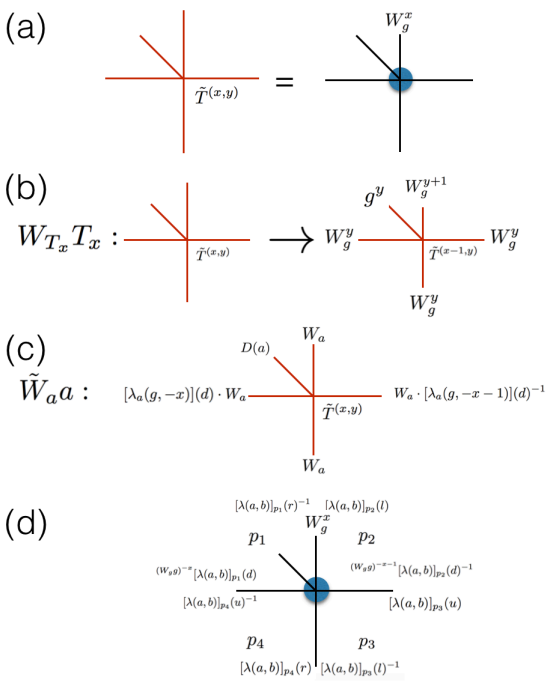


Figure 2.14: (a) The definition of the new tensor $\tilde{T}^{(x,y)}$ after the insertion of $[W_g(u)]^x$ to the upper leg of every original tensor $T^{(x,y)}$. (b) The new translation operation $W_{T_x}T_x$. It can be readily checked that $\tilde{T}^{(x,y)}$ is invariant under such translation. Note that we have $T_x = g^y T_x^{orig.}$, $T_y = T_y^{orig.}$ and $W_{T_y} = \mathbf{1}$. (c) The new on-site symmetry operation $\tilde{W}_a a$. It is shown in Fig. 2.15 that $\tilde{T}^{(x,y)}$ is invariant under such symmetry operation. (d) The new plaquette IGG for the new tensor $\tilde{T}^{(x,y)}$. As before, λ's from different plaquettes commute with each other, and the action of any two λ's in the same plaquette leave the tensor invariant. The tensor $\tilde{T}^{(x,y)}$ invariance under plaquette IGGs follows trivially from Fig. 2.13.

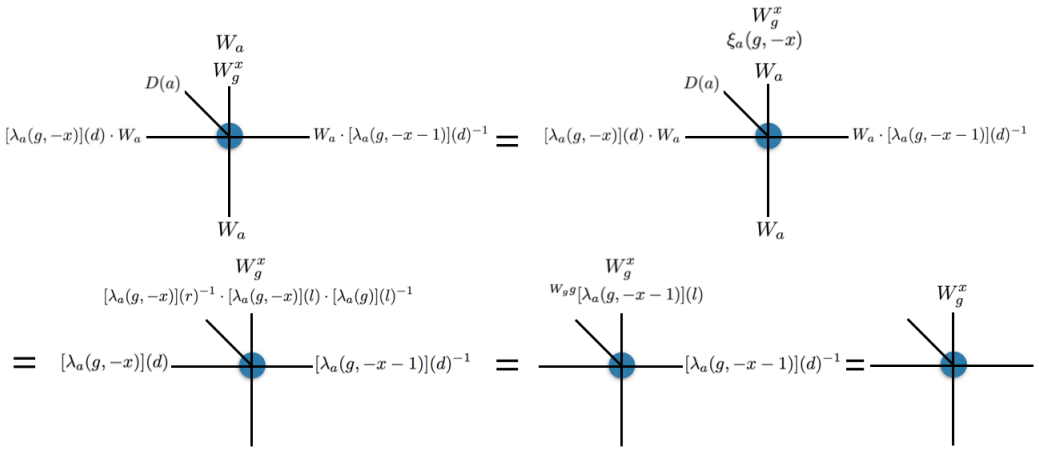


Figure 2.15: The revised tensor $\tilde{T}(x,y)$ is invariant under the newly-defined symmetry operation. The first equality comes from the commutation relation $W_a W_g^x = W_g^x \xi_a(g,-x) W_a$. In the second equality we have used the invariance of tensor under W_a as shown in Fig. 2.13 and the decomposition of $\xi_a(g,-x)$. In the third equality we have used the identity $\lambda_a(g,-x)(l) =^{W_g g} [\lambda_a(g,-x-1)](l)\cdot\lambda_a(g)(l)$. And we have also used the invariance of tensor under plaquette IGG as in Fig. 2.14 in the third and fourth equalities.

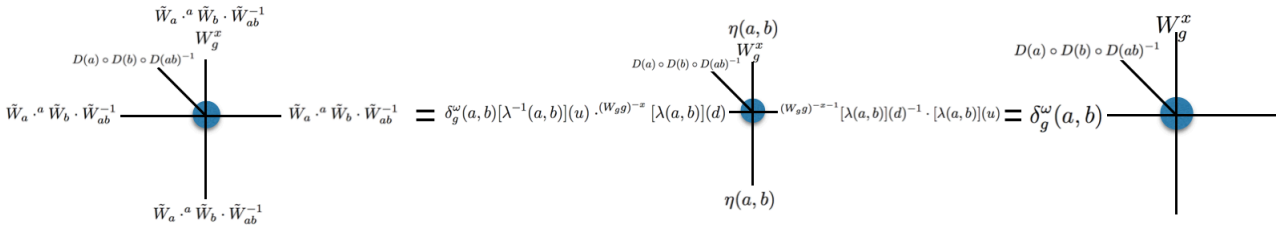


Figure 2.16: Every site-tensor carries a projective representation characterized by $[\delta_g^\omega(a,b)]^{-1}$. We show this by acting $\tilde{W}_a a \tilde{W}_b b (\tilde{W}_{ab} ab)^{-1}$ on both the physical legs and the virtual legs of tensor $\tilde{T}^{(x,y)}$, which should leave the tensor invariant without generating any phase. But from the calculation we find that the action on vitrual legs will contribute a factor $\delta_g^\omega(a,b)$, therefore the representation on the physical legs are $D(a) \cdot D(b) = [\delta_g^\omega(a,b)]^{-1} D(ab)$, *i.e.*, they are projected onto the $D_2(a)$ sector with projective representation. In the calculation above, we have used Eq. (2.95) in the first equality. And we have used the invariance of the tensor under the plaquette IGG defined in Fig. 2.14 in the second equality.

Bibliography

[1] E. Lieb, T. Schultz, and D. Mattis, Annals of Physics **16**, 407 (1961).

[2] M. Oshikawa, Phys. Rev. Lett. **84**, 1535 (2000).

[3] A. Paramekanti and A. Vishwanath, Phys. Rev. B **70**, 245118 (2004).

[4] M. B. Hastings, Physical Review B **69**, 104431 (2004).

[5] M. B. Hastings, EPL (Europhysics Letters) **70**, 824 (2005).

[6] M. Zaletel, A. Vishwanath, Physical Review Letters, **114(7)**, 077201 (2015).

[7] M. Cheng, M. Zaletel, M. Barkeshli, A. Vishwanath, and P. Bonderson, arXiv preprint arXiv:1511.02263 (2015).

[8] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).

[9] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[10] I. Affleck et al. Physical Review B **87**, 155114 (1987).

[11] X. Chen, Z.-C. Gu, Z.-X. Liu, and X.-G. Wen, Science **338**, 1604 (2012),

[12] X. Chen, Z.-C. Gu, Z.-X. Liu, and X.-G. Wen, Phys. Rev. B **87**, 155114 (2013).

[13] I. Dana, Y. Avron, and J. Zak, Journal of Physics C: Solid State Physics **18**, L679 (1985).

[14] Y.-M. Lu, Y. Ran, and M. Oshikawa, Annals of Physics **413**, 168060 (2020).

[15] J. Wu, T.-L. Ho, and Y.-M. Lu, arXiv preprint arXiv:1703.04776 (2017).

[16] D.-H. Lee and M. P. Fisher, Physical Review Letters **63**, 903 (1989).

[17] Y.-M. Lu and A. Vishwanath, Phys. Rev. B **93**, 155121 (2016).

[18] T. Senthil and M. Levin, Phys. Rev. Lett. **110**, 046801 (2013).

[19] L. Balents, M. P. Fisher, and C. Nayak, Physical Review B **60**, 1654 (1999).

[20] X.-G. Wen, Quantum

eld theory of many-body systems: from the origin of sound to an origin of light and electrons (Oxford University Press on Demand, 2004).

[21] S. Jiang and Y. Ran, Physical Review B **95**, 125107 (2017).

[22] K. Duivenvoorden, M. Iqbal, J. Haegeman, F. Verstraete, and N. Schuch, arXiv preprint arXiv:1702.08469 (2017).

[23] L. Balents, M. P. A. Fisher, and S. M. Girvin, Phys. Rev. B **65**, 224412 (2002).

[24] D. Sheng and L. Balents, Physical Review Letters **94**, 146805 (2005).

[25] S. Isakov, Y. B. Kim, and A. Paramekanti, Physical review letters **97**, 207204 (2006).

[26] E. Fradkin, Field theories of condensed matter physics (Cambridge University Press, 2013).

[27] D. S. Rokhsar and S. A. Kivelson, Physical review letters **61**, 2376 (1988).

[28] SPT phases protected by $G = SO(3) \times Z_2^{Ising}$ form a group $H^3(SO(3) \times Z_2^{Ising}, U(1))$. The Kunneth formula gives: $H^3(SO(3) \times Z_2^{Ising}, U(1)) = H^3(SO(3), U(1)) \times H^3(Z_2, U(1)) \times H^2(SO(3), Z_2) = Z \times Z_2 \times Z_2$. Following the Remark, it is straightforward to show that only the Z_2 index in $H^2(SO(3), Z_2)$ is enforced to be nontrivial. (Namely ω_0 in Theorem-(2) can be chosen to be the nontrivial element in $H^2(SO(3), Z_2)$,and the kernal $\mathcal{A}_g = H^3(SO(3), U(1)) \times H^2(SO(3), Z_2) = Z \times Z_2$.)

[29] Following the Kunneth formula: $H^3(Z_2^T \times Z_2^{Ising}, U(1)) = H^3(Z_2^T, U(1)) \times H^3(Z_2, U(1)) \times H^2(Z_2^T, Z_2) = Z_1 \times Z_2 \times Z_2$. In this example, the Z_2 index in $H^2(Z_2^T, Z_2)$ is enforced to be nontrivial, and the Z_2 index in $H^3(Z_2, U(1))$ is enforced to be trivial. This is because here ω_0 in Theorem-2 is the nontrivial element in $H^2(Z_2^T, Z_2)$, and the kernel subgroup $\mathcal{A}_g = Z_1$.

[30] X. Chen, Y.-M. Lu, and A. Vishwanath, Nature commu- nications **5** (2014).

[31] Now in this example, ω_0 in Theorem-2 is the nontrivial element in $H^3(Z_2, U(1))$, and the kernel subgroup $\mathcal{A}_g = Z_1$.

[32] R. Moessner and S. L. Sondhi, Phys. Rev. Lett. **86**, 1881 (2001).

[33] Y. Qi, Z.-C. Gu, and H. Yao, Physical Review B **92**, 155105 (2015).

[34] R. Moessner and S. Sondhi, Physical Review B **63**, 224401 (2001).

[35] R. Moessner, S. L. Sondhi, and E. Fradkin, Physical Re- view B **65**, 024504 (2001).

[36] R. Moessner, S. Sondhi, and P. Chandra, Physical review letters **84**, 4457 (2000).

[37] J. B. Kogut, Reviews of Modern Physics **51**, 659 (1979).

[38] A. Y. Kitaev, Annals of Physics **303**, 2 (2003).

[39] M. Levin and Z.-C. Gu, Phys. Rev. B **86**, 115109 (2012).

[40] D. Perez-Garcia et al. New J. Phys. **12**, 025010 (2010).

[41] H. Zhao, Z. Xie, Q. Chen, Z. Wei, J. Cai, and T. Xiang, Physical Review B **81**, 174411 (2010).

[42] S. Singh, R. N. C. Pfeifer, and G. Vidal, Phys. Rev. A **82**, 050301 (2010).

[43] S. Singh, R. N. Pfeifer, and G. Vidal, Physical Review B **83**, 115125 (2011).

[44] S. Singh and G. Vidal, Physical Review B **86**, 195114 (2012).

[45] B. Bauer et al. Physical Review B **83**, 125106 (2011).

[46] A. Weichselbaum, Annals of Physics **327**, 2972 (2012).

[47] S. Jiang and Y. Ran, Phys. Rev. B **92**, 104414 (2015).

[48] M. P. Zaletel, Physical Review B **90**, 235113 (2014).

[49] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, Phys. Rev. B **81**, 064439 (2010).

[50] X. Chen, Z.-C. Gu, and X.-G. Wen, Phys. Rev. B **83**, 035107 (2011).

Chapter 3

Divergent bulk photovoltaic effect in Weyl semimetals

3.1 Introduction

Weyl semimetals are a class of materials that can be regarded as three-dimensional analogs of grapheneupon breaking time-reversal or inversion symmetry. Electrons in a Weyl semimetal behave as Weylfermions, which have many exotic properties, such as chiral anomaly and magnetic monopoles in thecrystal momentum space. The surface state of a Weyl semimetal displays pairs of entangled Fermi arcs at two opposite surfaces. To linear order, the electrons near a Weyl node can be described by the effective two-by-two Hamiltonian: $H^{Weyl} = (\sum_a k^a \alpha^a - \mu)\sigma_0 + \sum_{a,b} k^a \beta_{ab} \sigma_b$, where μ is the chemical potential, $\vec{k}$ is the momentum relative to the Weyl node, $a/b = x, y, z$, σ_b are the Pauli matrices and σ_0 is the identity matrix. After choosing the frame of the principal axes and a proper basis, the

off-diagonal elements in β can be eliminated:

$$H^{Weyl} = (\hbar\vec{k}\cdot\vec{v}_t - \mu)\,\sigma_0 + \hbar\sum_a (k^a v^a \sigma_a). \tag{3.1}$$

Here apart from the generally anisotropic velocity $\vec{v}$, another velocity $\vec{v}_t$ describes the tilting of the bands and breaks the degeneracy between the bands at $\vec{k}$ and $-\vec{k}$, with the chemical potential (μ) usually nonzero.

It is known that either inversion symmetry or time-reversal symmetry needs to be broken in order to realize the Weyl nodes, and the experimentally confirmed Weyl semimetals have been overwhelmingly non-centrosymmetric and time-reversal symmetric, including the (Nb,Ta)(As,P) and $W_{1-x}Mo_xTe_2$ series [2–8]. When the tilting velocity is large enough, which is realized in materials like $W_{1-x}Mo_xTe_2$, the system can becomes a Type-II WSM [7], since along some directions the two energy bands share the same sign of the velocity. Such a system must have finite size Fermi surfaces (Fig.3.1(d))[1]. The type of a Weyl node is determined by a dimensionless number W:

$$W \equiv \sqrt{(v_t^x/v^x)^2 + (v_t^y/v^y)^2 + (v_t^z/v^z)^2}, \tag{3.2}$$

and a Type-I(Type-II) Weyl node is realized if $W < 1$ ($W > 1$).

As we show later, the bulk photovoltaic effect [10] with its direct connection to Berry curvature, [**?**, 11] offers a method to distinguish these two types. It has also attracted significant interest due to its potential applications in renewable energy generation [12] and fast photo-detectors [13, 14]. The intrinsic contributions to the BPVE can be expressed as

[1]Consequently it may be more accurate to call it a Type-II Weyl metal instead of a semimetal. However to be consistent with existing literature we follow the existing naming convention.

the second order nonlinear photocurrent response:

$$\mathbf{j}^a = \sigma_{\mathbf{2}}^{abc}(\omega)E^b(\omega)E^c(-\omega), \tag{3.3}$$

where, $a/b/c = x, y, z$, $\overrightarrow{\mathbf{j}}$ is the DC electric current density, $\vec{E}(t) = \mathrm{Re}[\vec{E}(\omega)e^{-i\omega t} + \vec{E}(-\omega)e^{i\omega t}]$ is the electric field of the light. In order to have a nonzero $\sigma_{\mathbf{2}}$, inversion symmetry needs to be broken, which happens to be also a condition to realize WSM. Generally speaking, $\sigma_{\mathbf{2}}^{abc}$ has both intraband and interband contributions. However, for time-reversal symmetric materials with linear polarized light, it turns out that this photocurrent response only has interband contributions, which has been coined the shift-current($\sigma_{\mathbf{2}}^{abc} = \sigma_{shift}^{abc}$) as it results from a change in the center of mass of the electrons upon optical excitation. Perturbation theory within the single-particle framework [11, 15, 16] gives:

$$\sigma_{shift}^{abc}(\omega) = \frac{2\pi e^3}{\hbar^2}\int \frac{d^3\vec{k}}{(2\pi)^3}\sum_{n,m} I_{mn}^{abc}[f_{nm}\cdot\delta(\omega_{mn}-\omega)], \tag{3.4}$$

where n, m label energy bands, $\hbar\omega_{mn}(\vec{k}) \equiv E_m(\vec{k}) - E_n(\vec{k})$ is the energy difference between the two bands. $f_{nm}(\vec{k}) = f_n(\vec{k}) - f_m(\vec{k})$ is the difference of the Fermi-Dirac function between the two bands. The gauge invariant quantity $I_{mn}^{abc}(\vec{k}) \equiv \frac{1}{2}\cdot\mathrm{Im}[r_{mn}^b r_{nm;a}^c + r_{mn}^c r_{nm;a}^b]$, where $r_{mn}^a(\vec{k}) \equiv i\langle u_m(\vec{k})|\partial_{k^a}|u_n(\vec{k})\rangle$ is nothing but the non-Abelian Berry's connection (with $r_{mn;b}^a(\vec{k})$ its generalized derivative: $r_{mn;b}^a \equiv \frac{\partial r_{mn}^a}{\partial k^b} - i[A_m^b(\vec{k}) - A_n^b(\vec{k})]r_{mn}^a(\vec{k})$, and $A_n^b(\vec{k}) \equiv i\langle u_n(\vec{k})|\partial_{k^b}|u_n(\vec{k})\rangle$ the usual intraband Berry's connection).

It has been pointed out that the shift current response is related to the topology of the band structure [**?**]. Indeed the quantities responsible for σ_{shift} directly involves the Abelian and non-Abelian Berry's phases. Because Weyl nodes are monopoles of the Abelian Berry's connection, these quantities are expected to be diverging and singular near Weyl nodes, which motivates us to carefully study the resulting nonlinear optical effects.

Eq.(3.4) has a familiar form of the Fermi's golden rule. Indeed, considering the case of linearly polarized light along the b-direction, one has:

$$I^{abb}_{mn} = |r^b_{nm}|^2 R^a_{nm,b}, \text{ (no summation on indices)} \tag{3.5}$$

where the gauge invariant real space displacement $\vec{R}_{nm,b}(\vec{k})$ is the so-called "shift-vector", defined as:

$$R^a_{nm,b}(\vec{k}) \equiv -\frac{\partial \mathrm{Arg}[r^b_{nm}]}{\partial k^a} + A^a_n(\vec{k}) - A^a_m(\vec{k}), \tag{3.6}$$

If one interprets Eq.(3.4,3.5) as the Fermi's golden rule, $|r^b_{nm}|^2$ is just the matrix-element factor for the optical absorption. Thus, $e\vec{R}_{nm,b}(\vec{k})$ should be viewed as the dipole moment induced the photo-excited particle-hole pair, giving rise to a rate of change in polarization, i.e., DC photocurrent.

Motivated by photogalvanic applications [12–14], previously the shift-current response has been mainly discussed in the context of insulators [11, 16], where the low temperature σ_{shift} vanishes when $\hbar\omega$ is below the band gap. It is convenient to introduce the optical joint density of states including the factor f_{nm} responsible for the Pauli-blocking effect:

$$JDOS(\omega) \equiv \int \frac{d^3\vec{k}}{(2\pi)^3} \sum_{m,n} [f_{nm} \cdot \delta(\omega_{mn} - \omega)] \tag{3.7}$$

For most materials, $I^{abc}_{mn}(\vec{k})$ is a smooth function of momentum, and according to Eq.(3.4) the response σ_{shift} is essentially proportional to $JDOS$. It has been proposed that engineering $JDOS$ in semiconductors may be a route to optimize the BPVE [17]. Although the $JDOS \propto \omega^2$ is small due to the linear dispersion in WSM, recent second harmonic [18] and photocurrent experiments [**?**, 19] show that these materials host large nonlinear optical effects at least in the infrared regimes. In this chapter we show that the WSM actually fea-

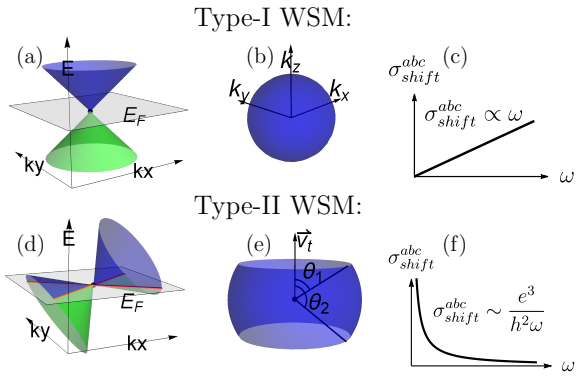


Figure 3.1: (color online) Considering the $\mu = 0$ situation, in (a),(d) we schematically plot the dispersion relations near a Type-I (Type-II) Weyl node. At zero temperature, the momentum space surfaces contributing to $JDOS$ (defined in Eq.(3.7)) are qualitatively different in (b) Type-I WSM and (e) Type-II WSM, leading to drastically different scaling behaviors of σ_{shift} shown in (c),(f).

ture divergent nonlinear optical responses in the low frequency regime due to the singular Berry's phases near the Weyl nodes.

3.2 Main results

Our main results are summarized in Fig. 3.1 and Fig. 3.2. For simplicity let us start with the zero doping case $\mu = 0$. In the absence of extrinsic scattering processes, simple dimensional analysis shows that up to fundamental constants $\frac{e^3}{h}$, $\sigma_{shift}(\omega)$ is proportional to the inverse of an energy scale, which could involve either an intrinsic energy scale t of the material, the temperature k_BT, or the photon energy $\hbar\omega$.

At zero doping and zero temperature, the linear Weyl equation Eq.(3.1) does not contain an energy scale, therefore $\sigma_{shift}(\omega)$ must be either $\sim \frac{e^3}{h^2}\frac{1}{\omega}$ or vanishing in this linear

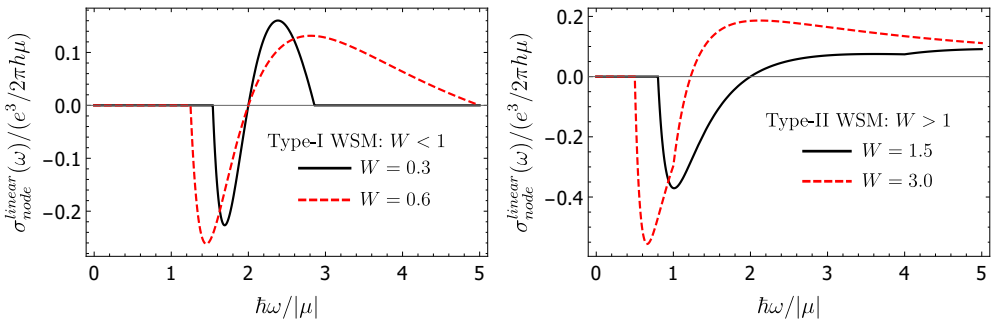


Figure 3.2: (color online) At $T = 0$, the quasi-universal (i.e., μ-independent) line shapes of the doping-induced peaks of σ_{shift} in Type-I (top) and Type-II(bottom) WSM based on results Eq.(3.9) of linearly dispersive nodes. The peak's frequency range has been re-scaled by a μ factor and its height has been re-scaled by a $1/\mu$ factor, the latter diverges in the low doping limit.

approximation. The former divergent case is exactly realized in Type-II WSM (Fig.3.1(f)). In Type-I WSM, however, we will show that zero doping and zero temperature leads to $\sigma_{shift}(\omega) \sim \frac{e^3}{h}\frac{\hbar\omega}{t^2}$ (Fig.3.1(c)), where t is an intrinsic energy scale due to the band-bending (i.e., deviation from linear dispersion) —typically ∼0.1-0.2eV.

Next, we discuss the reason for the drastic difference between Type-I and Type-II WSM, as well as the effect of doping. A careful evaluation of σ_{shift} in Eq.(3.4) based on the linear Weyl equation Eq.(3.1) shows that, at $T = 0$, a single Weyl node contributes (At $T \neq 0$, a general and more complicated analytic result is shown in supplemental material

Eq.(3.35,3.36)):

$$\sigma_{shift}^{abc} = \frac{\pi\chi}{8W^2} \cdot \sum_d \left[\epsilon_{dca} \frac{v_t^b v_t^d}{(v^d)^2} + (b \leftrightarrow c)\right] \cdot \sigma_{node}^{linear}(\omega) \tag{3.8}$$

where:

$$\sigma_{node}^{linear} \equiv (\cos[\theta_1]\sin[\theta_1]^2 - \cos[\theta_2]\sin[\theta_2]^2) \cdot \frac{e^3}{h^2\omega}, \tag{3.9}$$

$$\text{and: } \theta_1 \equiv \widetilde{\arccos}\left[\frac{\frac{2\mu}{\hbar\omega}+1}{W}\right],\ \theta_2 \equiv \widetilde{\arccos}\left[\frac{\frac{2\mu}{\hbar\omega}-1}{W}\right]. \tag{3.10}$$

Up to an order unity constant $\sigma_{shift}^{abc}(\omega)$ is determined by $\sigma_{node}^{linear}(\omega)$. Here W is defined in Eq.(3.2), ϵ_{abc} is the Levi-Civita antisymmetric tensor, and we have chosen the convention $\omega > 0$. $\chi \equiv \text{sign}[v^x v^y v^z] = \pm 1$ is the chirality of the Weyl node (i.e., monopole charge). Note that we have chosen the frame x, y, z to be right-handed. We have also defined a function $\widetilde{\arccos}[s] \equiv 0$ if $s \geqslant 1$, $\widetilde{\arccos}[s] \equiv \pi$ if $s \leqslant -1$, and $\widetilde{\arccos}[s] \equiv \arccos[s]$ if $-1 < s < 1$.

Interestingly, due to the chirality factor in this result, we know that this $1/\omega$ divergent term is absent in Dirac semimetals, because in those systems each Dirac node can be viewed as two Weyl nodes with opposite chiralities sharing the same set of μ, $\vec{v}_t$ and opposite $\vec{v}$. It can also be immediately seen from Eq. (3.8) that the $1/\omega$ divergence is absent for σ_{shift}^{aaa} due to the Levi-Civita antisymmetric tensor.

The physical meaning of the two angles θ_1,θ_2 is the following. For the moment it is convenient to re-scale the momentum in x, y, z directions so that $v^x = v^y = v^z = v$. The energy conservation $\delta(\omega_{mn} - \omega)$ constrains our consideration on a sphere in the re-scaled momentum space, whose radius equals $\frac{2\omega}{v}$. The re-scaled $\vec{v}_t$ can be used to define a special axis to set up a spherical coordinate system. Generally speaking, Pauli blocking takes over in certain solid angle regions. Namely the factor f_{nm} further constrains the sphere into a

region between the polar angles θ_1 and θ_2 (See Fig.3.1(e) for an illustration).

In Type-I WSM ($W < 1$) at zero doping, the whole sphere contributes, consistent with $\theta_1 = 0$ and $\theta_2 = \pi$ (Fig.3.1(b)). But this is exactly a situation when the θ-dependent factor in Eq.(3.9), and thus the $1/\omega$ term in σ_{shift}, vanishes. (Note that, in this case, a careful analysis including band-bending effects shows that even the constant order vanishes, leaving the next order $\sigma_{shift} \propto \omega$ in the low frequency limit. See supplemental information.) But in Type-II WSM ($W > 1$) at zero doping, only the part of the sphere between $\theta_{1,2} = \arccos[\pm\frac{1}{W}]$ contributes (Fig.3.1(e)), leading to the $1/\omega$ divergent response.

This results from the strongly angle dependent diverging Berry connection in WSM, despite the integrand I^{abc}_{mn} in Eq.(3.4) scaling as $\frac{1}{\omega^3}$. The full angular average over the 4π solid angle would annihilate the $1/\omega$ term in σ_{shift}. However, Pauli-blocking could take over in certain angular regions, removing the net cancellation and retaining the divergent term.

In particular, at $T = 0$, in Type-I WSM ($W < 1$) with finite μ, this $1/\omega$ term survives only over a frequency range: $\frac{2|\mu|}{1+W} \leqslant \hbar\omega \leqslant \frac{2|\mu|}{1-W}$. (Thus in the extreme case when $v_t = 0$ and $W = 0$, this term vanishes.) In fact, it is straightforward to show that, σ^{linear}_{node} would also change sign exactly at $\hbar\omega = 2|\mu|$. Altogether this leads to a large peak in the $\sigma_{shift}(\omega)$ with a width $\sim |\mu|/\hbar$ and a height $\sim \frac{e^3}{h} \cdot \frac{1}{|\mu|}$, featuring a characteristic sign-changing line-shape (Fig.3.2 top). Similarly a finite μ truncates the $1/\omega$ divergence in Type-II WSM and leads to a sign-changing large peak when $\hbar\omega \sim \mu$ (Fig.3.2 bottom).

As shown in Fig.3.2, the line-shapes of these peaks are quasi-universal: they only depend on the dimensionless number W. Different values of doping μ only re-scale the peak's frequency range by a μ factor and its height by a $1/\mu$ factor. This means that one may

use μ to control the frequency range of the peak to engineer tunable frequency-sensitive photo-electric devices.

Temperature and impurities— At finite temperatures, the Fermi-Dirac distribution would smear out and truncate the divergences when $\hbar\omega \lesssim k_B T$ (which is confirmed in our tight-binding model calculations Fig.3.3). But at low temperatures $k_B T \ll \hbar\omega$ these divergences are not significantly modified. (See supplemental material Eq.(3.35,3.36) for a general analytic form of the σ_{shift} within linear-dispersion approximation.) However even at zero temperature, impurities give rise to scattering, while finite temperature enables other scattering mechanisms (e.g. electron-phonon). These scatterings, which can be phenomenologically characterized by a scattering time τ, have been ignored so far. Namely, even at low temperatures $k_B T \ll \hbar\omega$ our result Eq.(3.8) holds only in the long scattering time $\omega\tau \gg 1$ limit with the divergences truncated when $\omega \lesssim \frac{1}{\tau}$. For instance, previous experiments report that τ in the Type-I WSM TaAs is of the order of a pico-second [20, 21]. This suggests that our predicted striking response can be observed in the Terahertz or higher frequency regimes.

A conceptually interesting situation occurs when a finite temperature is introduced in the $\mu = 0$ Type-I WSM, where divergence is absent at $T = 0$ due to angular cancellation discussed before. However when $T \neq 0$, the thermally excited particle-hole pairs partially play the role of doping, and there is no reason for a full angular cancellation. In this case, simple dimensional analysis leads to striking results: we expect a temperature-induced peak of the intrinsic σ_{shift} when $\hbar\omega \sim k_B T$, whose height $\sim \frac{e^3}{h k_B T}$ diverging in the low T limit. (see supplemental material Eq.(3.35) and Fig.3.4) This is observed in our tight-binding model calculations (see Fig.3.3(a)).

The size of the effects —It is interesting to estimate the size of effects in the divergent regimes; e.g., when $\hbar\omega \sim \mu$ for Type-I WSM or $\hbar\omega \gtrsim \mu$ in Type-II WSM, in the presence of a temperature $k_B T \ll \hbar\omega$. Previously shift-current responses have not be much studied in the low frequency (e.g. Terahertz) regimes. Note that even in a generic multiband metal, σ_{shift} is expected to vanish at zero temperature in the low frequency regimes. This is simply because the energy conservation $\delta(\omega_{mn} - \omega)$ and f_{nm} in $JDOS$ constrain both the valence band and the conduction band at the Fermi level when $\omega \to 0$, which would not occur due to band-repulsions.

Plugging in $\omega = 1$THz, the estimated size of $\sigma_{shift} \sim \frac{e^3}{\hbar^2}\frac{1}{\omega}$ in the divergent regimes is $\sim 0.01 A/V^2$, several orders of magnitudes larger than known reported values in visible or infrared regimes [22–27]

Tight-binding model— To concretely illustrate the predicted responses we compute the σ_{shift} tensor using a minimal time-reversal symmetric 4-band tight-binding model featuring 4 Weyl nodes: [18]

$$H^{TB} = t\big[(2.5 - \cos k^x - \cos k^y - \cos k^z)\tau_x + \sin k^y \tau_y$$
$$+ 0.5 \cos k^y s_x \tau_y + \sin k^z s_x \tau_z + (\xi \cos k^x - \mu_0)\big] - \mu, \tag{3.11}$$

where s_a(spin) and τ_a(orbital) are two sets of Pauli matrices. t is an overall energy scale which can be $0.1 - 0.5$eV, to be broadly consistent with the relevant energy scales in existing Weyl materials. Symmetries in this model include: time-reversal $is_y K$ (K is conjugation), $x \to -x$ mirror s_x, and $y \to -y$ mirror $s_y\tau_x$. The four Weyl nodes are located at $(k^x = \pm 0.920, k^y = \pm 0.464, k^z = 0)$, with an energy $\mu_0 t$ and $\mu_0 = 0.606\xi$. A positive parameter ξ controls the tilting velocity $\vec{v}_t$: for $\xi < 1 (\xi > 1)$, Type-I(Type-II) WSM is realized. We choose $\xi = 0.5$ ($\xi = 1.5$), corresponding to $W = 0.539 (W = 1.616)$,

as the representative for Type-I (Type-II) WSM.

We numerically compute σ^{abc}_{shift}, using the full formula Eq.(3.4) without resorting to the linear approximation. The results of a particular component σ^{zxx}_{shift} are plotted in Fig.3.3 for various doping levels and temperatures, which are fully consistent with previous discussions.

Second harmonic generation— Another directly related second order nonlinear optical response is the second harmonic generation (SHG), in which light at frequency ω drives current at frequency 2ω [28]. Defining SHG response tensor $\mathbf{j}^a(2\omega) = \sigma^{abc}_{SHG}(\omega, 2\omega)E^b(\omega)E^c(\omega)$, it is known that the real part $\mathrm{Re}[\sigma_{SHG}]$ is given by the interband contribution, and within the two-band approximation and linear-node approximation, we have

$$\mathrm{Re}[\sigma^{abc}_{SHG}(\omega, 2\omega)] = -\frac{3}{2}\sigma^{abc}_{shift}(\omega) \tag{3.12}$$

where $\sigma^{abc}_{shift}(\omega)$ given by Eq. (3.8) (see supplemental information). [2] [3]. For example, similar to σ_{shift}, the zero temperature $\mathrm{Re}[\sigma_{SHG}]$ features a similar $1/\omega$ divergence in Type-II WSM when $\mu = 0$, and large peak behaviors in both Type-I and Type-II WSM when $\mu \neq 0$. And even at $\mu = 0$, a finite temperature induces a large peak in $\mathrm{Re}[\sigma_{SHG}]$ in Type-I WSM with a width $\sim k_B T$ and a height $\sim \frac{e^3}{\hbar k_B T}$.

Possible applications— In this chapter we report the $1/\omega$ diverging DC photovoltaic effect in the low frequency regime in WSM, due to the combination of the diverging Berry's curvature and Pauli-blocking effect. Fast Terahertz photon detection has been a long

[2]Previously another identity is known within general two-band approximation: $\mathrm{Re}[\sigma^{aaa}_{SHG}(\omega, 2\omega)] = -\sigma^{aaa}_{shift}(\omega) + \frac{1}{2}\sigma^{aaa}_{shift}(2\omega)$ [**?**], which naively is inconsistent with Eq.(3.12). However within linear-node approximation we have shown that $\sigma^{aaa}_{shift}(\omega) = 0$, which is consistent with both identities.

[3]Interestingly, the intraband contribution of SHG is responsible for $\mathrm{Im}[\sigma_{SHG}]$, which has been pointed out to be a Fermi surface Berry's curvature effect and contains a $1/\omega$ term [31, 32]

standing challenge. The reported large photovoltaic effect may be useful for this purpose. In addition, the doping-induced large peak regimes of σ_{shift} in WSM, whose frequency ranges are controlled by μ, may be useful as a tunable frequency-sensitive probe for far-infrared or Terahertz photons, i.e., a spectrum analyzer.

3.3 Appendices

3.3.1 Shift current in type-I Weyl semi-metal

In this section we shall prove that for a generic type-I Weyl semi-metal with Fermi level at the Weyl nodes, the leading term in the shift current tensor will be proportional to ω when $\omega \to 0$.

The low energy physics can be captured by the following generic 2-band Hamiltonian with chemical potential μ set to zero

$$H = f_0\sigma_0 + \sum_{i=x,y,z} f_i\sigma_i, \tag{3.13}$$

where f_0, f_i are functions of k and $f_i = 0$ when $k = 0$. The eigenvalues are $E_c = f_0 + \epsilon$, $E_v = f_0 - \epsilon$ with $\epsilon = \sqrt{\sum_i f_i^2}$. Since the tilting will not affect the shift current tensor, we will set f_0 to be zero.

The shift current tensor for the 2-band model is obtained by doing the following integral [11, 17]

$$\sigma^{abc}_{shift}(\omega) = \frac{2\pi e^3}{\hbar^2}\int \frac{d^3\vec{k}}{(2\pi)^3} I^{abc}_{cv}[f_{cv}\cdot\delta(\omega_{cv}-\omega)], \tag{3.14}$$

where I^{abc}_{cv} has the following explicitly gauge-invariant expression

$$I^{abc}_{cv} = \sum_{i,j,m}[\frac{1}{8\epsilon^3}(f_m f_{i,b} f_{j,ac} - f_{i,b} f_{j,a} f_m \frac{\epsilon_{,c}}{\epsilon})\epsilon_{ijm} + (b \leftrightarrow c)], \tag{3.15}$$

where ϵ_{ijm} is the Levi-Civita symbol.

Now let's prove that terms proportional to $1/\omega$ (denoted as $\sigma^{(-1)}(\omega)$ below) and terms independent of ω (denoted as $\sigma^{(0)}(\omega)$ below) in σ^{abc} vanish in the low frequency limit $\omega \to 0$. A simple order of estimate with $k \sim \omega$ when $\omega \to 0$ tells us that only the k-linear terms and k-quadratic terms in f_i will contribute to $\sigma^{(-1)}(\omega)$ and $\sigma^{(0)}(\omega)$. Therefore it suffices to consider a linear node plus some quadratic corrections. It is always possible to choose the following form of f_i by an affine transformation which does not affect the integral ($\hbar v_F$ is set to be 1 throughout this section)

$$f_i = k_i + \alpha_{ijl} k_j k_l, \tag{3.16}$$

where α_{ijl} is a rank-3 tensor symmetric respect to the interchange $j \leftrightarrow l$ and repeated indices are summed over.

In the spherical coordinate system, we have

$$k_x = r\sin[\theta]\cos[\phi], k_y = r\sin[\theta]\sin[\phi], k_z = r\cos[\theta]. \tag{3.17}$$

And the unit vector $\hat{k} = (\sin[\theta]\cos[\phi], \sin[\theta]\sin[\phi], \cos[\theta])$ will be used below to simplify the notation.

When ω is small, $k_i \sim r \sim \omega$ and we can expand ϵ in powers of r

$$\epsilon = r + r^2 f(\theta, \phi) + \mathcal{O}(r^3), \tag{3.18}$$

where $f(\theta, \phi)$ is a function of $\alpha_{ijl}, \theta, \phi$. It is easy to obtain the explicit form of $f(\theta, \phi)$, but for our purpose we only need the fact that $f(\theta, \phi) = -f(\pi - \theta, \pi + \phi)$ since it comes from the angular dependence of terms involving 3 k_i's.

In the same spirit we can expand I^{abc} in powers of r

$$I^{abc} = \frac{1}{8\epsilon^3}[I^{(0)}_{abc} + I^{(1)}_{abc} + \mathcal{O}(r^2)], \text{ with}$$
$$I^{(0)} = -\epsilon_{bam}\frac{k_m k_c}{\epsilon^2} + (b \leftrightarrow c) = -\epsilon_{bam} r^2 \frac{\hat{k}_m \hat{k}_c}{\epsilon^2} + (b \leftrightarrow c),$$
$$I^{(1)} = 2\epsilon_{bjm}\alpha_{jac}k_m + \frac{h^{abc}(k_x, k_y, k_z)}{\epsilon^2} + (b \leftrightarrow c) = 2r\epsilon_{bjm}\alpha_{jac}\hat{k}_m + r^3\frac{h^{abc}(\hat{k}_x, \hat{k}_y, \hat{k}_z)}{\epsilon^2} + (b \leftrightarrow c) \tag{3.19}$$

where $h^{abc}(k_x, k_y, k_z)$ is a homogeneous polynomial of degree 3, whose explicit form, for our purpose, is not important.

The integral then becomes

$$\sigma^{abc}(\omega) = const * \int d\Omega (\int r^2 dr I^{abc}(r, \theta, \phi)\delta(2\epsilon(r, \theta, \phi) - \omega)), \tag{3.20}$$

where $d\Omega = \sin[\theta]d\theta d\phi$.

When $\omega \to 0$, we know that for any fixed θ, ϕ there is only one solution $r(\theta, \phi)$ to the equation $2\epsilon(r(\theta, \phi), \theta, \phi) = \omega$, from which we can solve r as a function of θ and ϕ. In fact, we can expand $r(\theta, \phi)$ in powers of ω with $f(\theta, \phi)$ defined before

$$r(\theta, \phi) = \frac{\omega}{2} - \frac{\omega^2}{4}f(\theta, \phi) + \mathcal{O}(\omega^3). \tag{3.21}$$

The integral can then be written as

$$\begin{aligned}
\sigma^{abc}(\omega) &= const * \int d\Omega \frac{1}{\omega^3}\frac{r(\theta, \phi)^2}{\partial_r \epsilon|_{r=r(\theta,\phi)}}[I^{(0)}_{abc} + I^{(1)}_{abc} + \mathcal{O}(\omega^2)] \\
&= const * \int d\Omega \frac{1}{\omega^3}[r(\theta, \phi)^2 - 2r(\theta, \phi)^3 f(\theta, \phi) + \mathcal{O}(\omega^4)][I^{(0)}_{abc} + I^{(1)}_{abc} + \mathcal{O}(\omega^2)] \\
&= const * \int d\Omega \frac{1}{\omega}[1 - 2\omega f(\theta, \phi) + \mathcal{O}(\omega^2)][I^{(0)}_{abc} + I^{(1)}_{abc} + \mathcal{O}(\omega^2)]
\end{aligned} \tag{3.22}$$

Terms that are of order $1/\omega$ comes from the integral over $I^{(0)}_{abc}$ with $r(\theta, \phi)$ set to $\omega/2$ in Eq.(3.22)

$$\sigma^{(-1)}(\omega) = const * \int d\Omega \frac{1}{\omega}(I^{(0)}_{abc}) = const * \int d\Omega \frac{1}{\omega}[-\epsilon_{bam}(\hat{k})_m(\hat{k})_c + (b \leftrightarrow c)] = const * (-\epsilon_{bac} - \epsilon_{cab}) = 0, \tag{3.23}$$

where we have used the fact that $\int(\hat{k})_m(\hat{k})_n d\Omega = \frac{4\pi}{3}\delta_{m,n}$.

Terms that are independent of ω comes from the following 3 integrals:

$$\sigma_I^{(0)} = const * \int d\Omega \frac{1}{\omega}(I_{abc}^{(0)}) = const * \int d\Omega f(\theta,\phi)[-\epsilon_{bam}\hat{k}_m\hat{k}_c + (b \leftrightarrow c)], \tag{3.24}$$

$$\sigma_{II}^{(0)} = const * \int d\Omega \frac{1}{\omega}(I_{abc}^{(1)}) = const * \int d\Omega[\epsilon_{bjm}\alpha_{jac}\hat{k}_m + h^{abc}(\hat{k}_x,\hat{k}_y,\hat{k}_z)/2 + (b \leftrightarrow c)] \tag{3.25}$$

and

$$\sigma_{III}^{(0)} = const * \int d\Omega f(\theta,\phi)(I_{abc}^{(0)}) = const * \int d\Omega f(\theta,\phi)[-\epsilon_{bam}(\hat{k})_m(\hat{k})_c + (b \leftrightarrow c)]. \tag{3.26}$$

It is easy to see that under $(\theta,\phi) \to (\pi-\theta,\pi+\phi)$, the integrand of Eq.(3.24),(3.25),(3.26) all change sign, therefore they are all equal to zero.

In conclusion, we have analytically shown that in the low frequency limit $\omega \to 0$, term that diverges as $1/\omega$ and term that is independent of ω in the shift current tensor σ^{abc} vanish. Therefore the leading term in σ^{abc} will be proportional to ω.

3.3.2 Analytical formula for the shift current in Weyl semi-metal with tilting and doping in low-frequency limit

In this section we will obtain an analytical formula of the shift-current tensor for a Weyl node with both tilting and nonzero chemical potential within the linear approximation at zero temperature. This formula captures the physics of type-I and type-II Weyl semimetals with or without doping in a unified fashion.

Let's consider the following generic Hamiltonian

$$H = (\hbar\vec{k}\cdot\vec{v}^t - \mu)\sigma_0 + \sum_{i=x,y,z}\hbar k_i v_i \sigma_i. \tag{3.27}$$

We will choose the following parameterization $v_i k_i = r O_{ij} \hat{k}_j$, where O_{ij} is an orthogonal matrix and $\hat{k} = (\sin[\theta]\cos[\phi], \sin[\theta]\sin[\phi], \cos[\theta])$. The orghogonal matrix O is chosen such that

$$(\frac{v_x^t}{v_x}, \frac{v_y^t}{v_y}, \frac{v_z^t}{v_z}) \cdot O = (0, 0, W), \tag{3.28}$$

where $W = \sqrt{(\frac{v_x^t}{v_x})^2 + (\frac{v_y^t}{v_y})^2 + (\frac{v_z^t}{v_z})^2}$. In fact the third column of O is fully determined

$$(O_{13}, O_{23}, O_{33}) = \frac{1}{W}(\frac{v_x^t}{v_x}, \frac{v_y^t}{v_y}, \frac{v_z^t}{v_z}). \tag{3.29}$$

In this new coordinate system, we have

$$\sigma_{shift}^{abc}(\omega) = \frac{2\pi e^3}{\hbar^2} \int \frac{d^3\vec{k}}{(2\pi)^3} I_{12}^{abc}[f_{21} \cdot \delta(\omega_{12} - \omega)] = \frac{e^3 \text{sign}[v_x v_y v_z]}{4\pi^2 \hbar^2 v_x v_y v_z} \int r^2 dr d\Omega I_{12}^{abc} f_{21} \cdot \delta(2r - \omega). \tag{3.30}$$

where I_{12}^{abc} is given by

$$-\frac{v_a v_b v_c}{8r^3}[O_{mj}\hat{k}_j O_{cl}\hat{k}_l \epsilon_{bam} + (b \leftrightarrow c)] \tag{3.31}$$

and f_{21} is given by

$$f_{21} = \frac{1}{e^{\beta(\hbar W r \cos[\theta] - \hbar r - \mu)} + 1} - \frac{1}{e^{\beta(\hbar W r \cos[\theta] + \hbar r - \mu)} + 1}, \tag{3.32}$$

which is a function of r and θ.

Therefore we have

$$\sigma_{shift}^{abc}(\omega) = -\frac{e^3 \text{sign}(v_x v_y v_z) v_a v_b v_c}{8\omega h^2 v_x v_y v_z} \int \sin[\theta] d\theta d\phi f_{21}(\omega/2, \theta)[O_{mj}\hat{k}_j O_{cl}\hat{k}_l \epsilon_{bam} + (b \leftrightarrow c)]. \tag{3.33}$$

The angular integration can be easily done:

$$\int d\Omega f_{21}(\omega/2, \theta)\hat{k}_m \hat{k}_n = \int d\Omega f_{21}(\omega/2, \theta)(\hat{k}_x^2 \delta_{m,n} + (\hat{k}_z^2 - \hat{k}_x^2)\delta_{m,n}\delta_{m,3}) = c \cdot \delta_{m,n} - \pi \cdot g(\varpi, z, W)\delta_{m,n}\delta_{m,3}, \tag{3.34}$$

where c is a constant, $z = e^{\beta\mu}$ is the fugacity and $\varpi = \frac{\beta\hbar\omega}{2}$. The isotropic part $c \cdot \delta_{m,n}$ does not contribute to the shift current tensor as in the case of type-I Weyl semimetal without doping. And the anisotropic part can be evaluated by first integrating over ϕ and then integrating over $x = \cos[\theta]$ to yield

$$g(\varpi, z, W) = -\int_{-1}^{1}(3x^2-1)f_{21}(\omega/2, x)dx = \frac{2}{W\varpi}\ln[\frac{1+z^{-1}e^{(1-W)\varpi}}{1+z^{-1}e^{(1+W)\varpi}} \cdot \frac{1+z^{-1}e^{(W-1)\varpi}}{1+z^{-1}e^{-(1+W)\varpi}}]$$
$$-\sum_{n=2,3}\frac{6}{W^n\varpi^n}[\mathrm{Li}_n(-z^{-1}e^{(1-W)\varpi}) + \mathrm{Li}_n(-z^{-1}e^{(1+W)\varpi}) - \mathrm{Li}_n(-z^{-1}e^{-(1+W)\varpi}) - \mathrm{Li}_n(-z^{-1}e^{(-1+W)\varpi})] \tag{3.35}$$

where $\mathrm{Li}_n(x)$ is the polylogarithm of order n.

Therefore after inserting Eq.(3.29) and Eq.(3.34) into Eq.(3.33) we have the following result

$$\sigma^{abc}_{shift}(\omega) = \frac{\pi \mathrm{sign}(v_x v_y v_z)}{8W^2} \cdot \left[\epsilon_{dca}\frac{v^t_b v^t_d}{(v_d)^2} + (b \leftrightarrow c)\right] \cdot \sigma^{linear}_{node}(\omega) \tag{3.36}$$
$$\text{where: } \sigma^{linear}_{node}(\omega) \equiv \left[g(\varpi, z, W) \cdot \frac{e^3}{h^2\omega}\right]$$

Therefore up to an order 1 constant, the shift-current response tensor is determined by the function σ^{linear}_{node}. Let's discuss two simple limits of Eq. (3.35). First, when $\mu = 0$, the function g only depends on the dimensionless variable ϖ, which gives us the scaling form

$$\sigma_{shift}(\omega)|_\beta = b \cdot \sigma_{shift}(b\omega)|_{\beta/b} \tag{3.37}$$

We plot this scaling behavior at finite temperatures and $\mu = 0$ for a few representative values of W in Fig.3.4. In particular for Type-I WSM at $\mu = 0$, $T \neq 0$ induces a peak of σ_{shift} whose width$\propto k_B T$ and height$\propto 1/k_B T$. This behavior is also observed in Fig.3(a) in our tight-binding model calculations.

Next, when $\beta \to \infty$, we can obtain a simpler form of $g(\varpi, z, W)$. In fact, an easier way is to replace the Fermi-Dirac distribution function by the Heaviside step function:

$$f_{21} = \Theta(r - \vec{k}\cdot\vec{v}^t + \mu/\hbar) - \Theta(-r - \vec{k}\cdot\vec{v}^t + \mu/\hbar) = \Theta(r(1 - W\cos[\theta]) + \mu/\hbar) - \Theta(-r(1 + W\cos[\theta]) + \mu/\hbar). \tag{3.38}$$

The Heaviside step function together with the δ function constraint the integration region of θ, therefore we will introduce the following two θ angles to characterize the upper and lower limit of the integration

$$\theta_1 \equiv \widetilde{\arccos}\big[\frac{\frac{2\mu}{\hbar\omega} + 1}{W}\big],\ \theta_2 \equiv \widetilde{\arccos}\big[\frac{\frac{2\mu}{\hbar\omega} - 1}{W}\big] \tag{3.39}$$

where the function $\widetilde{\arccos}[s]$ is defined in the following way: $\widetilde{\arccos}[s] \equiv 0$ if $s \geqslant 1$, $\widetilde{\arccos}[s] \equiv \pi$ if $s \leqslant -1$, and $\widetilde{\arccos}[s] \equiv \arccos[s]$ if $-1 < s < 1$.

After integration we have

$$\sigma^{linear}_{node}(\varpi, z, W)|_{T\to 0} = (\cos[\theta_1]\sin[\theta_1]^2 - \cos[\theta_2]\sin[\theta_2]^2)\frac{e^3}{h^2\omega}, \tag{3.40}$$

which gives us the shift-current tensor at zero-temperature:

$$\sigma^{abc}_{shift}(\omega) = \frac{\pi \mathrm{sign}(v_x v_y v_z)}{8W^2}\cdot\big[\epsilon_{dca}\frac{v^t_b v^t_d}{(v_d)^2} + (b \leftrightarrow c)\big]\cdot(\cos[\theta_1]\sin[\theta_1]^2 - \cos[\theta_2]\sin[\theta_2]^2)\frac{e^3}{h^2\omega} \tag{3.41}$$

3.3.3 Analytical formula for the second-harmonic-generation in Weyl semi-metal with tilting and doping in low-frequency limit

The second-harmonic-generation (SHG) response tensor is defined via $\mathbf{j}^a(2\omega) = \sigma^{abc}_{SHG}(\omega, 2\omega)E^b(\omega)E^c$

From the standard time-dependent perturbation theory [28,33,34], we have the following

expression for the real part of the SHG response tensor

$$\mathrm{Re}[\sigma^{abc}_{SHG}(\omega,2\omega)] = \frac{i\pi e^3}{2\hbar^2\omega^2}\sum_{m,n,p}\int_{BZ}\frac{dk^3}{(2\pi)^3}[(v^a_{mn}w^{bc}_{nm}+\frac{2v^a_{mn}\{v^b_{np}v^c_{pm}\}}{\omega_{mp}+\omega_{np}})f_{mn}\delta(2\omega-\omega_{nm})$$

$$+(w^{ab}_{mn}v^c_{nm}+w^{ac}_{mn}v^b_{nm})f_{mn}\delta(\omega-\omega_{nm})+\frac{v^a_{mn}\{v^b_{np}v^c_{pm}\}}{\omega_{pm}+\omega_{pn}}(f_{mp}\delta(\omega-\omega_{pm})-f_{np}\delta(\omega-\omega_{np}))], \tag{3.42}$$

where $v^i_{mn}\equiv\frac{1}{\hbar}\langle m|\,\partial_{k_i}H\,|n\rangle$, $w^{ij}_{mn}\equiv\frac{1}{\hbar^2}\langle m|\,\partial_{k_i}\partial_{k_j}H\,|n\rangle$ and $\{v^b_{np}v^c_{pm}\}=v^b_{np}v^c_{pm}+v^c_{np}v^b_{pm}$. Note that since we are dealing with generic tight-binding models, a careful derivation following Ref. [33] yields the extra w^{ij}_{mn} terms which are absent in the literatures listed above.

For a 2-band model within the linear approximation, Eq. (3.42) can be simplified to be

$$\mathrm{Re}[\sigma^{abc}_{SHG}(\omega,2\omega)] = \frac{i\pi e^3}{2\hbar^2}\int_{BZ}\frac{dk^3}{(2\pi)^3}\frac{v^a_{vc}\{v^b_{cv}\Delta^c_{cv}\}}{\omega^3_{cv}}f_{cv}[8\delta(2\omega-\omega_{cv})-\delta(\omega-\omega_{cv})], \tag{3.43}$$

where $\Delta^i_{nm}\equiv v^i_{nn}-v^i_{mm}$ and v/c corresponds to valence/conduction bands respectively.

In the linear approximation, we have the following identity

$$r^b_{mn}r^c_{nm;a}+r^c_{mn}r^b_{nm;a}=-\frac{v^a_{nm}\{v^b_{mn}\Delta^c_{mn}\}}{\omega^3_{mn}}-\frac{\Delta^a_{mn}\{v^b_{mn}v^c_{nm}\}}{\omega^3_{mn}}. \tag{3.44}$$

Therefore we can rewrite Eq. (3.43) as follows

$$\mathrm{Re}[\sigma^{abc}_{SHG}(\omega,2\omega)] = \frac{i\pi e^3}{2\hbar^2}\int_{BZ}\frac{dk^3}{(2\pi)^3}f_{vc}(r^b_{cv}r^c_{vc;a}+r^c_{cv}r^b_{vc;a})[8\delta(2\omega-\omega_{cv})-\delta(\omega-\omega_{cv})], \tag{3.45}$$

where we have used the fact that the integration over $\frac{\Delta^a_{cv}\{v^b_{cv}v^c_{vc}\}}{\omega^3_{cv}}$ vanishes due to time-reversal symmetry.

Eq. (3.45) assumes a very similar form to the expression of shift-current. In fact, it's easy to see that within the linear approximation we have $\mathrm{Re}[\sigma^{abc}_{SHG}(\omega,2\omega)]=-4\sigma^{abc}_{shift}(2\omega)+\frac{1}{2}\sigma^{abc}_{shift}(\omega)=-\frac{3}{2}\sigma^{abc}_{shift}(\omega)$, where $\sigma^{abc}_{shift}(\omega)$ is given by Eq. (3.36). As a consequence, all the

discussions of leading-order terms in

$$\mathrm{Re}[\sigma^{abc}_{SHG}(\omega, 2\omega)]$$

with the presence of tilting and doping at non-zero temperature naturally follow that of $\sigma^{abc}_{shift}(\omega)$ in the last section.

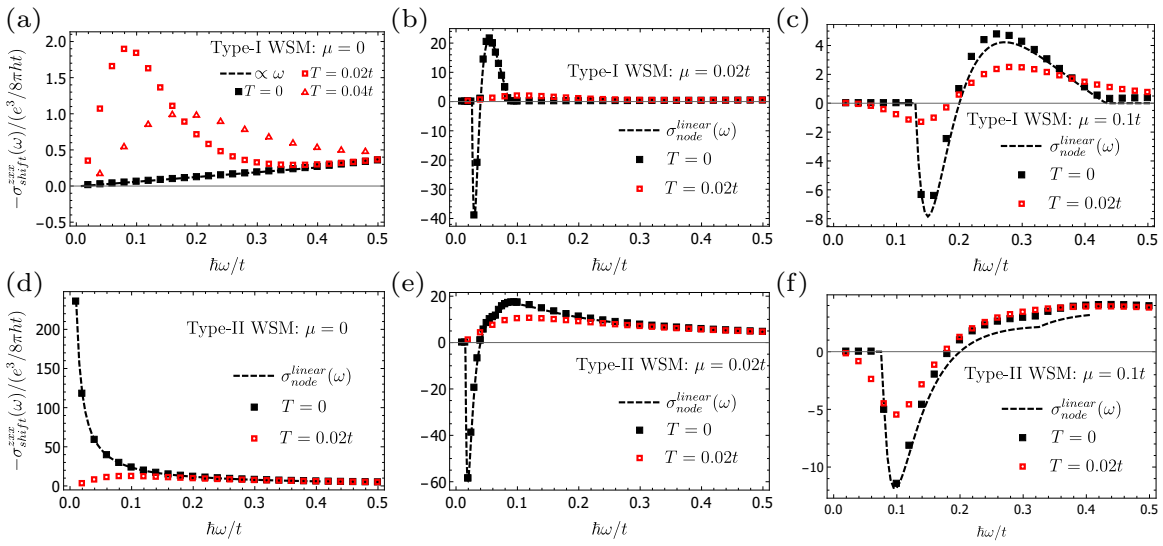


Figure 3.3: (color online) Numerically computed $\sigma^{zxx}_{shift}(\omega)$ using the full tight-binding model Eq.(3.11) with parameters in the main text (squares and triangles), comparing with analytic linear-node results after summing over four Weyl nodes $\sigma^{linear}_{node}(\omega)$ (dashed lines). At zero doping, (a): $\sigma_{shift} \propto \omega$ at $T = 0$ in Type-I WSM in the low frequency regime; a finite temperature partially plays the role of doping and induces a peak of σ_{shift} whose width $\propto T$ and height $\propto 1/T$ (see supplemental material Fig.3.4); (d): $\sigma_{shift} \propto 1/\omega$ at $T = 0$ in Type-II WSM, fully consistent with the result Eq.(3.8) within the linear approximation. This divergence is truncated by a finite temperature below $\hbar\omega \sim 5k_BT$. (b)(c)(e)(f): At finite dopings σ_{shift} feature large peaks whose width $\propto \mu$ and height $\propto 1/\mu$. At $T = 0$ these large peaks are well captured by Eq.(3.8) (the slight deviations for $\mu = 0.1t$ cases are due to expected band-bending effects.). At $k_BT = 0.02t$ the peaks for $\mu = 0.02t$ cases are strongly smeared out, while those for $\mu = 0.1t$ are quantitatively reduced.

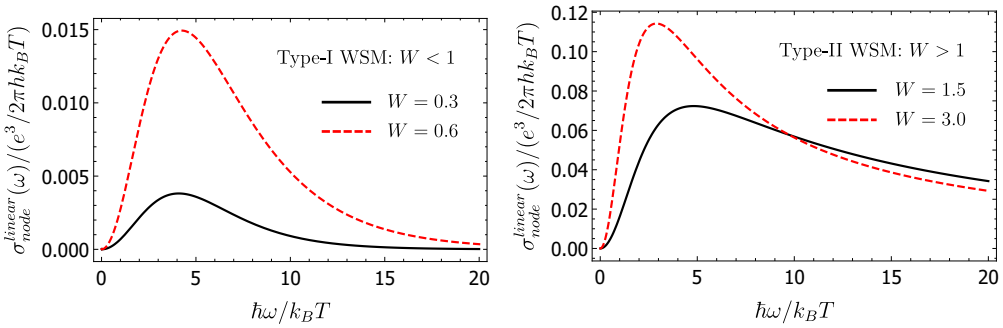


Figure 3.4: At zero doping $\mu = 0$, based on the linear-node result Eq.(3.35), we find that a finite temperature induces a peak of σ_{shift} in Type-I WSM (left), and truncate the $1/\omega$ divergence in Type-II WSM(right) when $\hbar\omega \sim k_B T$. Note that the frequency range is re-scaled by a $k_B T$ factor while σ_{shift} is re-scaled by a $1/k_B T$ factor. The line shapes of these curves only depend on W but independent of $k_B T$.

Bibliography

[1] N. Armitage et al., Reviews of Modern Physics, **90(1)**, 015001 (2018).

[2] B. Lv, H. Weng, B. Fu, X. Wang, H. Miao, J. Ma, P. Richard, X. Huang, L. Zhao, G. Chen, et al., Physical Review X **5**, 031013 (2015).

[3] S.-M. Huang, S.-Y. Xu, I. Belopolski, C.-C. Lee, G. Chang, B. Wang, N. Alidoust, G. Bian, M. Neupane, C. Zhang, et al., Nature communications **6**, 7373 (2015).

[4] S.-Y. Xu, I. Belopolski, N. Alidoust, M. Neupane, G. Bian, C. Zhang, R. Sankar, G. Chang, Z. Yuan, C.-C. Lee, et al., Science **349**, 613 (2015), ISSN 0036-8075.

[5] B. Lv, N. Xu, H. Weng, J. Ma, P. Richard, X. Huang, L. Zhao, G. Chen, C. Matt, F. Bisti, et al., Nat. Phys. **11**, 724 (2015).

[6] L. Yang, Z. Liu, Y. Sun, H. Peng, H. Yang, T. Zhang, B. Zhou, Y. Zhang, Y. Guo, M. Rahn, et al., Nat. Phys. **11**, 728 (2015).

[7] A. A. Soluyanov, D. Gresch, Z. Wang, Q. Wu, M. Troyer, X. Dai, and B. A. Bernevig, Nature **527**, 495 (2015).

[8] S.-Y. Xu, I. Belopolski, D. S. Sanchez, C. Zhang, G. Chang, C. Guo, G. Bian, Z. Yuan, H. Lu, T.-R. Chang, et al., Science advances **1**, e1501092 (2015).

[9] Note1,consequently it may be more accurate to call it a Type-II Weyl metal instead of a semimetal. However to be consistent with existing literature we follow the existing naming convention.

[10] P. Sturman, *Photovoltaic and Photo-refractive Effects in Noncentrosymmetric Materials*, vol. 8 (CRC Press, 1992).

[11] J. Sipe and A. Shkrebtii, Physical Review B **61**, 5337 (2000).

[12] K. Butler et al., Energy & Environmental Science **8**, 838 (2015).

[13] Q. Wang, J. Lai, and D. Sun, Optical Materials Express **6**, 2313 (2016).

[14] S. Young, in SPIE Defense Security (International Society for Optics and Photonics, 2016), pp. 98350P

[15] R. von Baltz and W. Kraut, Physical Review B **23**, 5590 (1981).

[16] C. Aversa and J. Sipe, Physical Review B **52**, 14636 (1995).

[17] A. M. Cook, B. M Fregoso, F. de Juan, S. Coh, and J. E. Moore, Nat. Comm **8**, 14176 (2017).

[18] L.Wu, S. Patankar, T.Morimoto, N. L. Nair, E. Thewalt, A. Little, J. G. Analytis, J. E. Moore, and J. Orenstein, Nat. Phys. **13**, 350 (2017).

[19] Q. Ma, S.-Y. Xu, C.-K. Chan, C. Zhang, G. Chang, Y. Lin, W. Xie, T. Palacios, H. Lin, S. Jia, et al., Nat. Phys. **13**, 842 (2017).

[20] S. Parameswaran, T. Grover, D. Abanin, D. Pesin, and A. Vishwanath, Physical Review X **4**, 031035 (2014).

[21] J. Behrends, A. G. Grushin, T. Ojanen, and J. H. Bardarson, Physical Review B **93**, 075114 (2016).

[22] D. Auston, A. Glass, and A. Ballman, Physical Review Letters **28**, 897 (1972).

[23] A. Glass, D. von der Linde, and T. Negran, Applied Physics Letters **25** (1974).

[24] W. Koch, R. Munser, W. Ruppel, and P. Wurfel, Ferroelectrics **13**, 305 (1976).

[25] A. Zenkevich, Y. Matveyev, K. Maksimova, R. Gaynutdinov, A. Tolstikhina, and V. Fridkin, Phys. Rev. B **90**, 161409 (2014).

[26] C. Somma, K. Reimann, C. Flytzanis, T. Elsaesser, and M. Woerner, Physical review letters **112**, 146602 (2014).

[27] F. Zheng, H. Takenaka, F. Wang, N. Z. Koocher, and A. M. Rappe, The journal of physical chemistry letters **6**, 31 (2014).

[28] R. W. Boyd, *Nonlinear optics* (Academic press, 2003).

[29] Note2, previously another identity is known within general two-band approximation: $\mathrm{Re}[\sigma^{aaa}_{SHG}(\omega,2\omega)] = -\sigma^{aaa}_{shift}(\omega) + \frac{1}{2}\sigma^{aaa}_{shift}(2\omega)$ [**?**], which naively is inconsistent with Eq.(3.12). However within linear-node approximation we have shown that $\sigma^{aaa}_{shift}(\omega) = 0$, which is consistent with both identities.

[30] Note3, interestingly, the intraband contribution of SHG is responsible for $\mathrm{Im}[\sigma_{SHG}]$, which has been pointed out to be a Fermi surface Berry's curvature effect and contains a $1/\omega$ term.

[31] I. Sodemann and L. Fu, Physical review letters **115**, 216806 (2015).

[32] T. Morimoto, S. Zhong, J. Orenstein, and J. E. Moore, Physical Review B **94**, 245121 (2016).

[33] D. Moss, E. Ghahramani, J. Sipe, and H. Van Driel, Physical Review B **41**, 1542 (1990).

[34] E. Ghahramani, D. Moss, and J. Sipe, Physical Review B **43**, 8990 (1991).

Chapter 4

Nonreciprocal directional dichroism induced by a temperature gradient as a probe for mobile spin dynamics in quantum magnets

4.1 Introduction

Quantum spin liquids(QSL), proposed by Anderson [1] for spatial dimensions > 1, attracted considerable interest in the past decades (see Ref. [2–5] for reviews). Although theoretically these novel states of matter are known to exist and have even been successfully classified [6, 7], to date there is no experimentally confirmed QSL material. As a matter of fact, an increasing list of candidate QSL materials emerges recently due to the extensive experimental efforts, including, for instance, Herbertsmithite [8, 9], α-$RuCl_3$ under a mag-

netic field [10], and quantum spin ice materials [11, 12]. An outstanding challenge in this field is the lack of appropriate experimental probes. Traditional probes for magnetic excitations include thermodynamic measurements, various spectroscopy measurements such as neutron scattering and nuclear magnetic resonance, and the thermal transport. Ideally, one would like to directly probe the mobile magnetic excitations in a QSL, such as the fractionalized spinons. The major limitation of traditional probes is from the contributions of other degrees of freedom; e.g., the spectroscopy measurements couple to local impurity modes, while the thermal transport couple to phonons. It is highly nontrivial to directly probe the intrinsic contribution from the mobile magnetic excitations. To highlight this challenge, there is no known direct probe to even detect the mobility gap of magnetic excitations, which is fundamentally important in the field of topologically ordered states.

In this chapter we propose a thermo-optic experiment which serves as a new probe for mobile magnetic excitations in Mott insulators respecting either the spatial inversion symmetry $\mathcal{I}$ or the time-reversal symmetry $\mathcal{T}$ [1], or both: the temperature-gradient-induced nonreciprocal directional dichroism (TNDD). In a sense TNDD combines the thermal transport and optical spectroscopy together, and effectively decouples from phonon and local magnetic modes.

4.2 The effect of TNDD

Theory of TNDD Nonreciprocal directional dichroism (NDD) is a phenomenon referring to the difference in the optical absorption coefficient between counterpropagating lights [14].

[1] or time-reversal symmetry combined with a spatial translation such as in an antiferromagnet

From the Fermi's golden rule, NDD for linearly polarized lights is due to the interference between the electric dipole and magnetic dipole processes [15][2]:

$$\delta_{\hat{n}}\alpha(\omega) \equiv \alpha_{\hat{n}}(\omega) - \alpha_{-\hat{n}}(\omega) = \frac{2\mu_r}{\epsilon_0 c^2}\frac{2\pi}{\hbar}\cdot\frac{\hbar\omega}{V}\sum_{i,f}(\rho_i - \rho_f)\cdot 2$$

$$\cdot \mathrm{Re}[\langle i|\vec{P}\cdot\hat{\mathcal{E}}|f\rangle\langle f|\vec{M}\cdot\hat{\mathcal{B}}|i\rangle]\cdot\delta(E_f - E_i - \hbar\omega) \tag{4.1}$$

where $\alpha_{\pm\hat{n}}(\omega)$ is the optical absorption coefficient of counterpropagating lights (along $\pm\hat{n}$) at frequency ω, which are $\mathcal{I}$ (or $\mathcal{T}$) images of each other. $\vec{P}$ ($\vec{M}$) is the electric polarization (magnetic moment) operator. $\hat{\mathcal{E}}$ ($\hat{\mathcal{B}}$) is the direction of the electric field (magnetic field) and $\hat{n} \sim \hat{\mathcal{E}} \times \hat{\mathcal{B}}$. $|i\rangle, |f\rangle$ label the initial and final states in the optical transition (ρ_i and ρ_f are their density matrix elements), ϵ_0 and c are the vacuum permittivity and the speed of light, V is the volume of the material, and μ_r is the material's relative permeability. Clearly both $\mathcal{I}$ and $\mathcal{T}$ need to be broken to have a nonzero NDD because $\mathrm{Re}[\langle i|P|f\rangle\cdot\langle f|M|i\rangle]$ is odd under either symmetry operation. NDD has been actively applied in the field of multiferroics [17–25] to probe the dynamical coupling between electricity and magnetism.

The TNDD spectroscopy essentially detects the joint density of states of mobile magnetic excitations, and can be intuitively understood as follows (see Fig.4.1(a)). Consider a Mott insulator respecting $\mathcal{I}$ and/or $\mathcal{T}$ so that NDD vanishes in thermal equilibrium. In the presence of a temperature gradient, the system reaches a nonequilibrium steady state with a nonzero heat current carried by mobile excitations. For simplicity one may assume that excitations of the system are well-described by quasiparticles, e.g., spinons or magnons, phonons, etc. The leading order nonequilibrium change of ρ_i and ρ_f in Eq.(4.1) satisfies

[2]In general NDD receives contributions from higher order multipole processes. [66] However in the context of Mott insulators the electric-dipole-magnetic-dipole contribution Eq.4.1 dominates.

$\delta\rho_i, \delta\rho_f \propto \nabla T \cdot \tau$ from a simple Boltzmann equation analysis, where τ is the relaxation time.

The crucial observation is that this nonequilibrium state breaks both the inversion symmetry (by ∇T) and the time-reversal symmetry (by τ). Consequently one expects a NDD signal proportional to $\nabla T \cdot \tau$. Precisely speaking TNDD is a second-order thermo-electromagnetic nonlinear response: it is a change of optical absorption (a linear response) due to a temperature gradient. The factor $\nabla T \cdot \tau$ in TNDD indicates that it is a generalization of Drude-phenomenon to nonlinear responses. Notice that the Drude-phenomenon is independent of whether the system has a quasiparticle description or not. Even in the absence of quasiparticle descriptions, strongly interacting liquids may have nearly conserved momentum. The relaxation time τ in Drude physics should be interpreted as the momentum relaxation time [26]. This indicates that TNDD discussed here can be generalized to systems without quasiparticle descriptions such as the U(1)-Dirac spin liquid [27–29] and the spinon Fermi surface state [30, 31].

Advantages of TNDD spectroscopy Now we comment on the major advantages of TNDD as a probe of spin dynamics. First, TNDD is a dynamical spectroscopy with the frequency resolution in contrast to the DC thermal transport, and essentially probes the joint density of states of magnetic excitations. Second, the fact that TNDD only receives contributions from $\mathrm{Re}[\langle i|P|f\rangle \cdot \langle f|M|i\rangle]$ dictates that the *phonons*' contribution can be safely ignored: The natural unit for the magnetic moment of phonon, the nuclear magneton, is more than three orders of magnitudes smaller than that of the electron, the Bohr magneton.

In addition, at the intuitive level, a local magnetic mode (e.g. from a magnetic impurity

atom) can only couple to a local temperature instead of a temperature gradient. A local temperature respects both $\mathcal{I}$ (after taking disorder-average) and $\mathcal{T}$. Consequently, such local modes are not expected to contribute to TNDD either. From a more careful estimate (see App.4.4.1 for detailed discussions), we find that the contribution to TNDD from localized modes with a localization length ξ, comparing to the contribution from the intrinsic mobile magnetic modes, is at least down by a factor of ξ/l_m, where l_m is the mean-free path of the mobile magnetic excitations. We have assumed that $\xi \ll l_m$: for local magnetic modes carried by magnetic impurity atoms or crystalline defects, typically ξ is comparable with the lattice spacing a, while usually $l_m \gg a$ in a reasonably clean Mott insulator at low temperatures.

Estimate of the TNDD response One may estimate the size of TNDD signal in a spin-orbital coupled Mott insulator. The relevant dimensionless ratio limiting the experimental resolution is:

$$TNDD(\omega) \equiv \frac{\delta_{\hat{n}}\alpha(\omega)}{\alpha_{\hat{n}}(\omega) + \alpha_{-\hat{n}}(\omega)}. \tag{4.2}$$

In a Mott insulator, the polarization carried by a magnetic excitation can be estimated as $\zeta \cdot e \cdot a$, where a is the lattice spacing and ζ is dimensionless. Assuming the average temperature of the system $k_B T$ to be comparable to the magnetic excitation energy[3], we find that (see App.4.4.2 for details):

$$TNDD(\omega) \sim \left(\frac{D}{J}\right)^2 \frac{\zeta}{\boldsymbol{\alpha}} \cdot \frac{|\nabla T| \cdot l_m}{T} \sim \left(\frac{D}{J}\right)^2 \cdot \frac{|\nabla T| \cdot l_m}{T}, \tag{4.3}$$

[3]Similar to a thermal transport experiment, if the temperature of the system is far below the magnetic excitation energy, a temperature gradient would not efficiently affect the magnetic excitation distributions and would not lead to a sizable TNDD.

in the limit of a weak spin-orbit coupling. Here $\boldsymbol{\alpha} \approx 1/137$ is the fine-structure constant and we used $\zeta \sim 10^{-2} \sim \boldsymbol{\alpha}$ in typical transition metal Mott insulators [33]. Notice that in the absence of spin-orbit coupling, TNDD vanishes since the spin magnetic moment M is a spin-triplet[4]. D and J are the Dzyaloshinskii-Moriya(DM) interaction and the exchange interaction respectively. In a system with a strong spin-orbit coupling one may set $D/J \sim 1$, and $TNDD(\omega)$ is proportional to the ratio of the temperature change across l_m and the temperature. To optimize signal, one may choose a large temperature gradient such that $\nabla T \cdot \boldsymbol{w} \sim T$ where $\boldsymbol{w}$ is the linear system size along the ∇T direction, and $TNDD(\omega) \sim l_m/\boldsymbol{w}$. For instance, l_m of magnetic excitations in a quantum spin ice material was reported to be of the order of a micron [35]. For a typical millimeter sample size, $TNDD(\omega)$ can be as large as 10^{-3}, well detectable within the currently available experimental technology.

Crystal symmetry analysis TNDD can be phenomenologically described by a tensor η:

$$\delta_{\hat{n}}\alpha(\omega) = \sum_{a,b,c} \eta_{abc}(\omega)\hat{\mathcal{E}}_a\hat{\mathcal{B}}_b\nabla_c T \tag{4.4}$$

The symmetry condition for $\eta_{abc}(\omega)$ is determined by the fusion rule of two vectors $(\hat{\mathcal{E}}, \nabla T)$ and one pseudovector $(\hat{\mathcal{B}})$ into a trivial representation under the point group. For any point group, symmetry always allows nonzero η_{abc}: one may always consider the case $\hat{n} \sim \hat{\mathcal{E}} \times \hat{\mathcal{B}}$ to be parallel to ∇T.

As an example, we find that there are four independent response coefficients for the D_{3d}

[4]We only consider the contribution from the spin magnetic moment in this chapter. The orbital magnetic moment in a Mott insulator is a spin-singlet but is much smaller than the spin magnetic moment, by a factor of $(t/U)^2$ in the (t/U)-expansion. [33, 67]

point group:

$$\delta_{\hat{n}}\alpha = \eta_1 \nabla_z T(\hat{\mathcal{E}} \times \hat{\mathcal{B}})_z + \eta_2 \hat{\mathcal{B}}_z(\hat{\mathcal{E}} \times \nabla T)_z + \eta_3 \hat{\mathcal{E}}_z(\hat{\mathcal{B}} \times \nabla T)_z$$
$$+ \eta_4 \big[(\hat{\mathcal{E}}_x \hat{\mathcal{B}}_y + \hat{\mathcal{E}}_y \hat{\mathcal{B}}_x)\nabla_y T - (\hat{\mathcal{E}}_x \hat{\mathcal{B}}_x - \hat{\mathcal{E}}_y \hat{\mathcal{B}}_y)\nabla_x T\big] \tag{4.5}$$

Here the x-axis is a C_2-axis and the yz-plane is a σ_d mirror-plane in the D_{3d} group. The D_{3d} point group is realized in the QSL candidate Herbertsmithite, in the Heisenberg model on the Kagome lattice with DM interactions (see below and Fig.4.1(b)), as well as in the generalized Kitaev-Heisenberg model on the honeycomb lattice [36–38], relevant for Na_2IrO_3 [39] and $RuCl_3$ [40–42].

Microscopic model We present a concrete microscopic calculation for the TNDD spectrum. The nearest neighbor spin-1/2 Hamiltonian under consideration is on the kagome lattice:

$$H = J \sum_{<ij>} \vec{S}_i \cdot \vec{S}_j + \sum_{<ij>} \vec{D}_{ij} \cdot \vec{S}_i \times \vec{S}_j, \tag{4.6}$$

This model is relevant for various QSL candidate materials such as $ZnCu_3(OH)_6Cl_2$ (Herbertsmithite) and $Cu_3Zn(OH)_6FBr$, and respects both $\mathcal{T}$ and $\mathcal{I}$. Based on the D_{3d} crystal symmetry for the kagome plane, the DM vector $\vec{D}_{ij} = -\vec{D}_{ji}$ has two independent coupling constants: D_z (out-of-plane) and D_p (in-plane) [43] (see Fig.4.1(b)). Precisely speaking:

$$\vec{D}_{ij} = d_{ij} \cdot (D_z \cdot \hat{z} + D_p \cdot \hat{z} \times \hat{r}_{ij}), \tag{4.7}$$

where $d_{ij} = \pm 1$, $\hat{r}_{ij}$ is the unit vector along the direction from the site-j to the site-i. As shown in Fig.4.1(b), in each bow-tie: $d_{12} = d_{23} = d_{31} = 1, d_{34} = d_{45} = d_{53} = -1$.

Dipole-coupling with an external electric field $\delta H = -\vec{E} \cdot \vec{P}$, the electric polarization $\vec{P}$

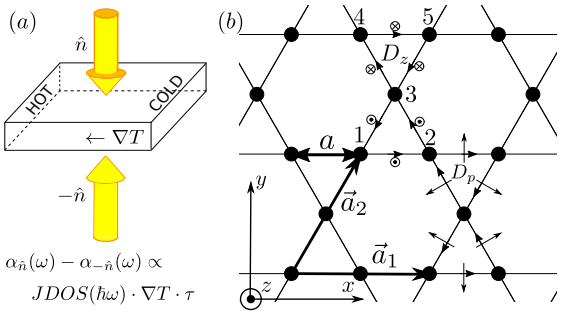


Figure 4.1: (a): A schematic illustration of the TNDD effect: in the presence of a temperature gradient, the optical absorption coefficients for counterpropagating lights become different, which essentially probes the joint density of states of mobile magnetic excitations. (b) A Kagome lattice and the Dzyaloshinskii–Moriya vectors $\vec{D}_{ij}$.

has the following form for the nearest neighbor terms [44][5]:

$$\begin{aligned} P_y =& \frac{\zeta e a}{\sqrt{3}}[\vec{S}_3 \cdot (\vec{S}_2 + \vec{S}_1 - \vec{S}_5 - \vec{S}_4) - 2\vec{S}_1 \cdot \vec{S}_2 + 2\vec{S}_5 \cdot \vec{S}_4], \\ P_x =& \zeta e a \cdot [\vec{S}_3 \cdot (\vec{S}_2 - \vec{S}_1 + \vec{S}_5 - \vec{S}_4)], \end{aligned} \tag{4.8}$$

where $e < 0$ is the electron charge, a is the nearest neighbor distance, and ζ is a dimensionless coupling constant (in this chapter $\vec{S} = \vec{\sigma}/2$.) ζ can be generated via a t/U expansion in a Hubbard model [46]. In the leading order $J = \frac{4t^2}{U}$ and $\zeta = \frac{12t^3}{U^3}$ [44]. [6]

$Q_1 = Q_2$ Z_2 spin liquid: Schwinger boson mean-field treatment There are extensive numerical evidences that the Heisenberg model on the kagome lattice may realize a QSL ground state, although the nature of the QSL is under debate [29, 48–52]. The

[5]Generally the polarization operator contains spin-triplet terms similar to DM interactions. Here for simplicity we only consider spin-singlet terms which dominate in the weak spin-orbit coupling limit.

[6]ζ also receives contribution from the magneto-elastic coupling. For a typical transition metal Mott insulator, this contribution to polarization is similar in size as the contribution from the t/U-expansion [33, 44].

present work does not attempt to resolve this long-standing puzzle. Instead, we will focus on one candidate spin liquid state, which may be realized in the model Eq.(4.6): Sachdev's $Q_1 = Q_2$ Z_2 QSL [53]. The $Q_1 = Q_2$ QSL is a gapped state and can be described using the Schwinger boson mean-field theory [54–56], in which spin is represented by bosonic spinons: $\vec{S}_i = \frac{1}{2} b^\dagger_{i\alpha} \vec{\sigma}_{\alpha\beta} b_{i\beta}$, while boson number per site is subject to the constraint $b^\dagger_{i\alpha} b_{i\alpha} = 2S$. We then do the usual mean-field decoupling and diagonalize the quadratic mean-field spinon Hamiltonian to obtain three spinon bands. We treat DM interaction as a perturbation and keep contributions up to the linear order of D/J. Under this approximation we arrive at the following mean-field Hamiltonian.

$$\begin{aligned} H_{MF} = &-\mu \sum_i (b^\dagger_{i\alpha} b_{i\alpha} - 2S) - \frac{J}{2} \sum_{\langle ij \rangle} (A^*_{ij} \hat{A}_{ij} + h.c.) \\ &+ \sum_{\langle ij \rangle} (\frac{\vec{D}_{ij}}{4} \cdot A_{ij} \hat{\vec{C}}^\dagger_{ij} + h.c.), \end{aligned} \tag{4.9}$$

where operators $\hat{A}_{ij} \equiv b_{i\alpha} \epsilon_{\alpha\beta} b_{j\beta}$ and $\hat{C}_{ij} \equiv -i b_{i\alpha} (\epsilon \vec{\sigma})_{\alpha\beta} b_{j\beta}$. H_{MF} may be viewed as an ansatz to construct variational spin-liquid wavefunctions with parameters A_{ij}, μ. In Sachdev's $Q_1 = Q_2$ state, A_{ij} have the following spatial pattern: $A_{ij} = d_{ij} A$, and A can be chosen to be real. See Appendix. 4.4.3 for more details.

After Bogoliubov diagonalization, three bands are found: $H_{MF} = \sum_{\vec{k}, u=1,2,3}^{\alpha=\uparrow,\downarrow} E_{u,\vec{k}} \gamma^{\alpha\dagger}_{u,\vec{k}} \gamma^{\alpha}_{u,\vec{k}}$ as shown in Fig.4.2, where $\uparrow, \downarrow$ label the Kramers degeneracy. Tuning chemical potential μ so that the band structure is near the boson condensation at Γ, the lowest energy band $u = 1$ is well described by a relativistic boson disperson:

$$E_{1,\vec{k}} \approx \sqrt{\Delta^2 + \hbar^2 k^2 v^2}, \tag{4.10}$$

where Δ is the bosonic spinon gap.

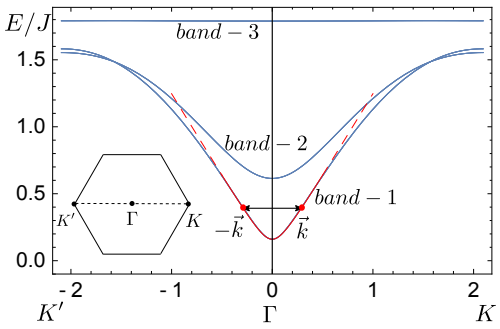


Figure 4.2: The Schwinger boson band dispersion (blue solid lines) for the mean-field Hamiltonian Eq.(4.9) of Sachdev's $Q_1 = Q_2\ Z_2$ QSL with parameters $A = 1$, $D_z = D_p = 0.1J$, and $\mu = -1.792J$. The low energy band-1 near the Γ point is well described by the relativistic dispersion Eq.(4.10) with gap $\Delta = 0.16J$ (red line). The two-spinon (red dots at $\pm\vec{k}$) contribution to the TNDD response computed in Eq.(4.11) and App.4.4.3 is illustrated.

TNDD contributed from the bosonic spinons In the low temperature limit, the two-spinon contribution dominates TNDD with $|f\rangle \sim \gamma^{\alpha,\dagger}_{u,\vec{k}}\gamma^{\beta\dagger}_{v,-\vec{k}}|i\rangle$ in Eq.(4.1). [7]. ρ_i, ρ_f in Eq.(4.1) is related to the nonequilibrium bosonic spinon occupation $g_{u,\vec{k}}$. From a simple Boltzmann equation analysis with a single relaxation time τ, $g_{u,\vec{k}}$ deviates from the equilibrium occupation $g^0_{u,\vec{k}} = \frac{1}{e^{\beta(\vec{r})E_{u,\vec{k}}}-1}$ by $\delta g_{u,\vec{k}} = \frac{\partial g^0_{u,\vec{k}}(\vec{r})}{\partial E}E_{u,\vec{k}}\frac{\tau\vec{v}_{u,\vec{k}}\cdot\nabla T}{T(\vec{r})}$, where $\vec{v}_{u,\vec{k}} = \frac{\partial E_{u,\vec{k}}}{\hbar\partial\vec{k}}$. This $\delta g_{u,\vec{k}}$ is responsible for TNDD.

Since TNDD is a bulk response we consider a 3D system consisting of stacked 2D layers each described by the model Eq.(4.6) with an interlayer distance d. Using the electric polarization Eq.(4.8) and spin magnetic moment $\vec{M} = g_s\mu_B\vec{S}$, in App.(4.4.3 we compute the low temperature/energy TNDD response tensor defined in Eq.(4.4) within our mean-

[7]Notice that a single spinon excitation is not gauge invariant and does not contribute to physical responses

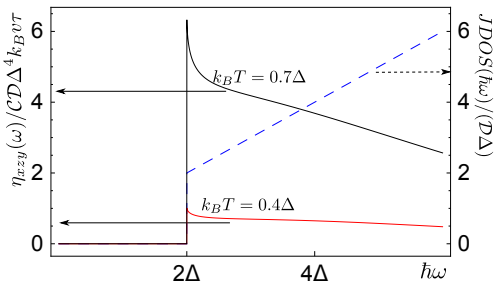


Figure 4.3: The bosonic two-spinon contribution to TNDD spectra of Sachdev's $Q_1 = Q_2$ Z_2 QSL Eq.(4.9) at the temperature $k_BT = 0.7\Delta$ (solid black line) and $k_BT = 0.4\Delta$ (solid red line), together with the two-spinon joint density of states (dashed blue line).

field treatment (corresponding to η_2 in Eq.(4.5)). As plotted in Fig.4.3, we find that (x, y, z-directions are illustrated in Fig.4.1)

$$\begin{aligned}\eta_{xzy}(\omega) &= \mathcal{C} \cdot \left[1 + 2g^0(\hbar\omega/2)\right] \cdot (k_BT)^3 \\ &\cdot \left[3G_3(z) - 3\ln z \cdot G_2(z) + (\ln z)^2 G_1(z)\right] \\ &\cdot e^{-\sqrt{(\hbar\omega/2)^2 - \Delta^2}/\Delta} \cdot \hbar\omega \cdot JDOS(\hbar\omega) \cdot \frac{\tau \cdot v}{T}. \end{aligned} \tag{4.11}$$

Here the constant $\mathcal{C} \equiv 8\pi \boldsymbol{\mathcal{u}}_0 \boldsymbol{\alpha}^2 \zeta a a_0 \cdot \frac{\mu_r g_s a^3}{\hbar^3 v^3}$, where a_0 is the Bohr radius. $\boldsymbol{\mathcal{u}}_0 \propto (D/J)^2$ is a dimensionless constant related to the mean-field band structure and can be determined numerically. For the parameters $D_z = D_p = 0.1J$ and $\mu = -1.792J$ we find that $\boldsymbol{\mathcal{u}}_0 = 0.603$. The 3D optical joint density of states $JDOS(\hbar\omega) \equiv \mathcal{D} \cdot \hbar\omega \cdot \Theta(\hbar\omega - 2\Delta)$ where $\mathcal{D} \equiv \frac{1}{8\pi\hbar^2 v^2 d}$. $g^0(\hbar\omega/2) = \frac{1}{e^{\hbar\omega/2k_BT}-1}$, $z \equiv e^{-\frac{\Delta}{k_BT}}$, and $G_\nu(z) \equiv \frac{1}{\Gamma(\nu)} \int_0^\infty \frac{x^{\nu-1}dx}{z^{-1}e^x - 1}$ is the Bose-Einstein integral. Eq.(4.11) holds when the temperature and the photon energy are within the regime of the relativistic dispersion Eq.(4.10).

In the limit $k_BT \ll \Delta$, Eq.(4.11) can be simplified and we have $\eta_{xzy}(\omega) \propto e^{-\Delta/k_BT}$, where the thermal activation factor can be traced back to $\delta g_{\vec{k}}$. Importantly, beyond the mean-field treatment, TNDD is generally $\propto \delta\rho_i, \delta\rho_f \propto \nabla T \cdot \tau$ in Eq.(4.1), and a thermal

activation factor $e^{-\Delta/k_BT}$ in TNDD is always due to the energy diffusion near the mobility gap Δ. Therefore TNDD can serve as a sharp measurement of the *mobility gap* Δ of the magnetic excitations.

4.3 Discussion and conclusion

Bosonic vs. fermionic spinons We computed the TNDD response contributed from bosonic spinons in the Sachdev's $Q_1 = Q_2$ Z_2 QSL. Fermionic spinons also exist in this Z_2 QSL and their contribution to TNDD can be similarly computed in a dual Abrikosov fermion approach [58,59]. Without pursuing this calculation in details, one expects that the bosonic factor $[1 + 2g^0(\hbar\omega/2)]$ (Bose-Einstein integrals) in Eq.(4.11) will be replaced by the corresponding fermionic factor $[1 - 2f^0(\hbar\omega/2)]$ (Fermi-Dirac integrals), where $f^0(\hbar\omega/2) = 1/(e^{\hbar\omega/2k_BT} + 1)$. The contributions from the bosonic spinons and fermionic spinons have different temperature dependence, which, in principle, may be used to detect the statistics of quasiparticles in certain situations.

Magnetically ordered states It is also interesting to consider the TNDD response in a conventional magnetically ordered state respecting either $\mathcal{I}$, or $\mathcal{T}$ combined with a lattice-translation symmetry (as in the case of an antiferromagnet), or both. One may similarly consider the two-magnon contribution to the TNDD response, which probes the joint density of states of magnons. Our estimate Eq.(4.3) will be modified as follows (see Appendix 4.4.2 for details). If the magnetic order is non-collinear, which breaks spin-rotational symmetry completely, the $(D/J)^2$ factor in Eq.(4.3) is replaced by ~ 1. If the magnetic order is collinear, which only breaks the spin-rotation symmetry down to $U(1)$, the $(D/J)^2$ factor

is replaced by D/J.

Conclusion In this chapter we propose the temperature-gradient-induced nonreciprocal directional dichroism (TNDD) spectroscopy experiment in Mott insulators. Comparing with traditional probes for magnetic excitations, TNND spectroscopy has unique advantages: it directly probes mobile magnetic excitations and decouples from local impurity modes and phonon modes. For instance, an activation behavior $\propto e^{-\Delta/k_BT}$ in the temperature dependence of TNDD sharply measures the *mobility gap* Δ of the magnetic excitations, a quantity challenging to measure using traditional probes but of fundamental importance in the field of topologically ordered QSL.

The present work can be viewed as one example in a large category of nonlinear thermo-electromagnetic responses. There are other interesting effects. For instance, a temperature gradient also induces a circular dichroism in a system respecting both $\mathcal{T}$ and $\mathcal{I}$. We leave these other responses as topics of future studies.

4.4 Appendices

4.4.1 Localized modes

Let us consider the situation of a Mott insulator in the presence of impurities/disorders, which could introduce localized magnetic modes. Below we consider the contribution to TNDD response from these localized modes.

Firstly, we comment on the meaning of "localized modes" discussed here. In an isolated localized phase of matter, like a many-body localized phase(see Ref. [60, 61] for reviews), thermalization breaks down and the meaning of a temperature-gradient is unclear. We are

NOT discussing the TNDD response in this situation.

In realistic quantum materials, the magnetic localized modes are coupled with a thermal bath (e.g., phonon thermal bath) and a local temperature is well defined. To facilitate the discussion, one may consider a system with a $U(1)$ spin rotation symmetry in order to sharply define a magnetic localized mode. In addition, we assume a finite mobility gap Δ of the $U(1)$ charge, and magnetic localized excitations may exist below Δ. Assuming l_m being the mean-free path for mobile magnetic excitations, practically the localized magnetic modes may fall into two regimes according to the localization length ξ:

(1): $\xi \ll l_m$. This is the more common situation realized in practical materials. Here the localized magnetic modes may be extrinsic magnetic impurity atoms, or may form at crystalline defects. They may also form at the centers of the vortices of valence bond solid (VBS) order [62]. Typically the localization length ξ of these magnetic modes is of the same order as the lattice spacing a, while $l_m \gg a$ in a reasonably clean Mott insulator.

It is difficult to model a magnetic localized mode with $\xi \sim a$ since lattice scale details cannot be neglected. Instead, we consider the following situation $a \ll \xi \ll l_m$ so that a low energy effective description is still valid. As a crude model for such magnetic localized modes, one may consider a quantum dot of size ξ in the presence of a temperature gradient; for instance, the left (right) edge of the quantum dot is in contact with a heat reservior at temperature T_L (T_R). The modes in the quantum dot are travelling ballistically since $\xi \ll l_m$. Consequently the right-mover (left-mover) in the quantum dot is at temperature T_L (T_R). Such a nonequilibrium ensemble is quantitatively comparable with a large (energy-)diffusive system in the presence of the same temperature gradient but with $l_m \sim \xi$ (for example, see Eq.(4.24)). Namely, in the present situation, ξ replaces the role of l_m in our

estimate Eq.(4.3). we conclude that the dimensionless ration $TNDD(\omega)$ contributed by such localized modes is reduced by a factor of $\sim \xi/l_m$.

(2): $\xi \gg l_m$**.** In this situation, the system hosts would-be mobile modes. These modes scatter with disorder multiple times before eventually become localized. For instance, Anderson weak-localization in two spatial dimensions happens with ξ parametrically larger than l_m. It is instructive to consider a system size L satisfying $\xi > L > l_m$. For such a system size the localization physics is not present yet. Because photon absorption is still a local process, we expect that the contribution to the TNDD response from such localized modes to be comparable with that from mobile modes.

In summary, the contribution to TNDD response from localized modes in the regime $\xi \ll l_m$ can be safely neglected. In the opposite regime $\xi \gg l_m$, the localized modes still contribute to TNDD significantly. Nevertheless, the localized modes in the latter regime are would-be extended (mobile) states in the absence of disorder.

4.4.2 Spin-orbit coupling and the estimate of TNDD response

From the discussion in the main text and Eq.(4.1), up to matrix element effects, the TNDD spectroscopy directly probes the joint density of states $JDOS(\hbar\omega)$ of the mobile magnetic excitations:

$$\delta_{\hat{n}}\alpha(\omega) \equiv \alpha_{\hat{n}}(\omega) - \alpha_{-\hat{n}}(\omega) \propto \hbar\omega \cdot JDOS(\hbar\omega) \cdot \nabla T \cdot \tau \tag{4.12}$$

In order to estimate the optical absorption coeffient $\alpha_{\hat{n}}$ in a Mott insulator, one need to estimate the strength of electric polarization and the magnetic dipole moment. It turns out that they are comparable in a typical transition metal Mott insulator, which is drastically different from the case of a band metal/insulator. In the latter case the electric polarization

carried by a typical particle-hole excitation is $\sim e \cdot a$ where e is the electron charge and a is the lattice constant, while the magnetic moment carried by the same excitation is of the order of a Bohr magneton μ_B. For a given electromagnetic wave, the magnetic dipole energy scale $\mu_B \cdot B$ is smaller than the electric dipole energy scale $e \cdot a \cdot E$ by roughly a factor of the fine-structure constant $\sim 1/137$, which is why the magnetic dipole processes are often neglected in a band metal/insulator.

In a Mott insulator, however, the electric polarization carried by a magnetic excitation is heavily reduced. In the framework of the Hubbard model, this electric polarization can be estimated as $\zeta \cdot e \cdot a$ where the dimensionless factor $\zeta \sim 8(t/U)^3$ [33]. On the other hand, the magnetic dipole moment carried by the same excitation is still $\sim \mu_B$. As a result, they would have comparable sizes for typical 3d transition metal Mott insulators with $t/U \sim 10$.

The absorption coefficient due to the electric dipole processes can be estimated based on the Fermi's golden rule:

$$\begin{aligned} \alpha_{\hat{n}}(\omega) &\sim \frac{2}{n_r \epsilon_0 c} \frac{2\pi}{\hbar} \cdot |\langle f|P|i\rangle|^2 \cdot \hbar\omega \cdot JDOS(\hbar\omega) \\ &\sim \frac{16\pi^2}{n_r} \boldsymbol{\alpha} \zeta^2 a^2 \cdot \hbar\omega \cdot JDOS(\hbar\omega). \end{aligned} \tag{4.13}$$

where n_r is the relative refractive index of the material, c is the speed of light, $\boldsymbol{\alpha}$ is the fine structure constant $\sim 1/137$, and $JDOS(\hbar\omega)$ is the joint density of states for the relevant excitations at photon energy $\hbar\omega$. We assume that the temperature is comparable with the magnetic excitation energy scale, and we have used the typical matrix element $\langle f|P|i\rangle \sim \zeta \cdot e \cdot a$ where a is the lattice constant.

Notice that $JDOS(\hbar\omega)$ may be estimated as $\sim \frac{1}{a^3}\frac{1}{W}$ where a is the lattice constant and W is the band width of the excitations. For a typical photon energy $\sim W$, one finds

that $\hbar\omega \cdot JDOS(\hbar\omega) \sim 1/a^3$, independent of the nature of the excitations. For instance, the interband absorption coefficient $\alpha(\omega)$ in a band metal/insulator is typically $\sim 10^7 \mathrm{m}^{-1}$. The dimensionless coupling constant ζ reduces by a factor of 10^2 in transition metal Mott insultors, which gives the absorption coefficient $\sim 10^3 \mathrm{m}^{-1}$, broadly consistent with the tera-Hertz penetration depth ($\sim$ 1mm) for these quantum magnets [63, 64].

The TNDD response can be similarly estimated. We first consider the case of a quantum paramagnet.

$$\delta_{\hat{n}}\alpha(\omega) \sim \frac{2\mu_r}{\epsilon_0 c^2}\frac{2\pi}{\hbar} \cdot (\rho_i - \rho_f) \cdot 2 \cdot \mathrm{Re}[\langle f|P|i\rangle\langle i|M|f\rangle] \cdot \hbar\omega \cdot JDOS(\hbar\omega) \tag{4.14}$$

We again assume that the temperature is comparable with the magnetic excitation energy scale, and consequently the effect of temperature gradient in $(\rho_i - \rho_f)$ can be estimated by the dimensionless factor $\frac{|\nabla T|\cdot l_m}{T}$ where l_m is the mean-free path of the magnetic excitations. *If the spin-orbit coupling (SOC) is strong* one may estimate $\langle f|P|i\rangle \sim \zeta e a$ while $\langle i|M|f\rangle \sim g_s\mu_B$ (g_s is the g-factor the spin magnetic moment.). Putting together we have:

$$\delta_{\hat{n}}\alpha(\omega) \sim 16\pi^2\mu_r g_s \boldsymbol{\alpha}^2 \zeta a_0 a \cdot \frac{|\nabla T| \cdot l_m}{T} \cdot \hbar\omega \cdot JDOS(\hbar\omega), \quad \text{if strong SOC.} \tag{4.15}$$

Here a_0 is the Bohr radius.

From Eq.(4.13,4.15), and $a_0 \sim a$, we can estimate that if the spin-orbit coupling is strong and the temperature is comparable with the magnetic excitation energy scale, the dimensionles ratio $TNDD$ in Eq.(4.2)

$$TNDD(\omega) \sim \frac{\boldsymbol{\alpha}}{\zeta}\frac{|\nabla T| \cdot l_m}{T} \sim \frac{|\nabla T| \cdot l_m}{T}, \text{ if strong SOC.} \tag{4.16}$$

Here we used the fact that for a typical transition metal Mott insulator $\zeta \sim 10^{-2} \sim \boldsymbol{\alpha}$.

In the absence of the SOC, $\langle i|M|f\rangle = 0$ because $\vec{M} = g_s\mu_B\vec{S}$ is proportional to the conserved total spin $\vec{S}$ (We only consider the spin magnetic moment. The orbital magnetic moment in Mott insulators is much smaller and neglected.). In the limit of a weak SOC: $D/J \ll 1$, the TNDD response can be estimated as follows. The only effect of the weak SOC is in the matrix element product: $\langle f|P|i\rangle\langle i|M|f\rangle$.

For the magnetic dipole matrix element: $\langle i|M|f\rangle \propto \frac{1}{E_f - E_i}\langle i|[S, H]|f\rangle \propto \langle f|\vec{D}\cdot[S, \vec{S}_i \times \vec{S}_j]|i\rangle$. Notice that the operator of the commutator is a spin triplet. There are two possibilities: (1): the states $|f\rangle$ and $|i\rangle$ differ by spin-1 in the limit $D/J \to 0$. For instance, $|f\rangle$ may be a spin triplet while $|i\rangle$ is a spin singlet in that limit; (2): the states $|f\rangle$ and $|i\rangle$ have the same spin in the limit $D/J \to 0$.

In the situation-(2), the magnetic dipole matrix element $\langle i|M|f\rangle \propto (D/J)^2$, because the wavefunction corrections of $|f\rangle$ and $|i\rangle$ due to nonzero D/J need to be considered. In this situation, the electric dipole matrix element $\langle f|P|i\rangle \propto (D/J)^0$ since P is a spin singlet operator in the limit of $D/J \to 0$. Therefore in situation-(2) we have $\langle f|P|i\rangle\langle i|M|f\rangle \propto (D/J)^2$.

In the situation-(1), a similar consideration leads to: $\langle i|M|f\rangle \propto (D/J)$ and $\langle f|P|i\rangle \propto (D/J)$. So we still have $\langle f|P|i\rangle\langle i|M|f\rangle \propto (D/J)^2$.

In summary, we have the following estimate in a quantum paramagnet assuming the temperature is comparable with the magnetic excitation energy scale:

$$\begin{aligned} TNDD(\omega) &\sim \left(\frac{D}{J}\right)^2 \frac{\boldsymbol{\alpha}}{\zeta}\frac{|\nabla T|\cdot l_m}{T} \\ &\sim \left(\frac{D}{J}\right)^2 \frac{|\nabla T|\cdot l_m}{T}, \text{ if weak SOC.} \end{aligned} \tag{4.17}$$

Next we estimate the TNDD response in magnetic ordered states due to magnon excitations. Even in the absence of microscopic SOC, the $(D/J)^2$ factor in the estimate Eq.(4.17) will be replaced by ~ 1 in a non-collinear magnetic ordered state, because the spin-rotation symmetry is completely broken.

In a collinear magnetic ordered state, the spin rotation symmetry is broken down to $U(1)$ in the absence of SOC. The electric polarization operator P is expected to carry zero charge under this $U(1)$ rotation. To have a nonzero matrix element product $\langle f|P|i\rangle\langle i|M|f\rangle$, one must consider the linear-order effect of the SOC. Therefore in this case the $(D/J)^2$ factor in the estimate Eq.(4.17) will be replaced by $\sim D/J$.

4.4.3 Details of the mean-field calculation for TNDD

In this section we provide a detailed account of the Schwinger boson mean-field theory. The spin is represented by bosonic spinons

$$\vec{S}_i = \frac{1}{2} b_{i\alpha}^\dagger \vec{\sigma}_{\alpha\beta} b_{i\beta}, \tag{4.18}$$

while boson number per site is subject to the constraint:

$$b_{i\alpha}^\dagger b_{i\alpha} = \kappa. \tag{4.19}$$

Although $\kappa = 2S$ for spin-S, it will be convenient to consider κ to be a continuous parameter, taking on any non-negative value [53, 65].

Considering the operator identities $\vec{S}_i \cdot \vec{S}_j = -\frac{1}{2}\hat{A}_{ij}^\dagger \hat{A}_{ij} + \frac{\kappa^2}{4}$ and $\vec{S}_i \times \vec{S}_j = \frac{1}{4}[\hat{\vec{C}}_{ij}^\dagger \hat{A}_{ij} + h.c.]$, where

$$\hat{A}_{ij} = -\hat{A}_{ji} = b_{i\alpha}\epsilon_{\alpha\beta} b_{j\beta}, \qquad \hat{\vec{C}}_{ij} = \hat{\vec{C}}_{ji} = -\mathrm{i} b_{i\alpha}(\epsilon\vec{\sigma})_{\alpha\beta} b_{j\beta}. \tag{4.20}$$

standard mean-field decoupling of Eq.(4.6) leads to the mean-field Hamiltonian:

$$\begin{aligned} H_{MF} = & -\frac{J}{2}\sum_{<ij>}(A_{ij}^{*}\hat{A}_{ij} + A_{ij}\hat{A}_{ij}^{\dagger} - |A_{ij}|^{2}) \\ & + \sum_{<ij>}\frac{\vec{D}_{ij}}{4}\cdot(\vec{C}_{ij}^{*}\hat{A}_{ij} + A_{ij}\hat{\vec{C}}_{ij}^{\dagger} - \vec{C}_{ij}^{*}A_{ij} + h.c.) \\ & - \mu\sum_{i}(b_{i\alpha}^{\dagger}b_{i\alpha} - \kappa). \end{aligned} \tag{4.21}$$

Here the chemical potential μ is introduced to enforce constraint Eq.(4.19) on the mean-field level. H_{MF} may be viewed as an ansatz to construct variational spin-liquid wavefunctions with parameters $A_{ij}, \vec{C}_{ij}, \mu$.

We will consider the case of a small D/J and keep contributions up to the linear order of D/J. Under this approximation we will set the parameter (*not the operator* $\hat{\vec{C}}_{ij}$) $\vec{C}_{ij} \propto D/J$ to zero in Eq.(4.21) below, which yields Eq.(4.9) in the main text. We also focus on Sachdev's $Q_1 = Q_2$ state, where A_{ij} happens to have the following spatial pattern: $A_{ij} = d_{ij}A$, and A is chosen to be real.

After diagonalizing H_{MF} in the momentum space, there are three Kramers degenerate Bogoliubov boson bands (see Fig.4.2):

$$H_{MF} = \sum_{\vec{k},\alpha=\uparrow,\downarrow}^{u=1,2,3} E_{u,\vec{k}}\gamma_{u,\vec{k}}^{\alpha\dagger}\gamma_{u,\vec{k}}^{\alpha}. \tag{4.22}$$

Notice that spin is not a good quantum number and $\uparrow, \downarrow$ are simply labelling the two-fold Kramers degeneracy for each band.

In the presence of a temperature gradient $\nabla T(\vec{r})$, the occupation of Bogoliubov spinons $g_{u,\vec{k}} = \langle n_{u,\vec{k}}\rangle$ (where $n_{u,\vec{k}} = \gamma_{u,\vec{k}}^{\dagger}\gamma_{u,\vec{k}}$) deviates from the thermal equilibrium value $g_{u,\vec{k}}^{0}$. For simplicity, we consider the steady state Boltzmann equation within a single relaxation-time

approximation:

$$\vec{v}_{u,\vec{k}} \cdot \nabla_{\vec{r}} g_{u,\vec{k}}(\vec{r}) = -\frac{g_{u,\vec{k}}(\vec{r}) - g^0_{u,\vec{k}}(\vec{r})}{\tau}, \tag{4.23}$$

where $g^0_{u,\vec{k}}(\vec{r}) = \frac{1}{e^{E_{u,\vec{k}}/k_B T(r)}+1}$, $\vec{v}_{u,\vec{k}} = \frac{1}{\hbar}\nabla_{\vec{k}} E_{u,\vec{k}}$. To the leading order, these give $\delta g_{u,\vec{k}}(\vec{r}) \equiv g_{u,\vec{k}}(\vec{r}) - g^0_{u,\vec{k}}(\vec{r})$:

$$\delta g_{u,\vec{k}}(\vec{r}) \equiv \delta g^{\uparrow}_{u,\vec{k}}(\vec{r}) = \delta g^{\downarrow}_{u,\vec{k}}(\vec{r}) = \frac{\partial g^0_{u,\vec{k}}(\vec{r})}{\partial E} E_{u,\vec{k}} \frac{\tau \vec{v}_{u,\vec{k}} \cdot \nabla T}{T(\vec{r})} \tag{4.24}$$

Since the velocity $\vec{v}_{u,\vec{k}} = -\vec{v}_{u,-\vec{k}}$, we have:

$$\delta g_{u,\vec{k}}(\vec{r}) = -\delta g_{u,-\vec{k}}(\vec{r}) \tag{4.25}$$

To be concrete, we focus on the case $\hat{\mathcal{E}} = \hat{x}$ and $\hat{\mathcal{B}} = \hat{z}$, with the light propagating direction $\hat{n} = -\hat{y}$ and the temperature gradient $\nabla T \propto \hat{y}$(the η_2 response in Eq.(4.5)). In order to compute the matrix elements in Eq.(4.1), one writes P_x and M_z in terms of the Bogoliubov bosons, and selects the relevant terms:

$$\begin{aligned}
P_x \to & \sum_{\vec{q}}^{v,\alpha,w,\alpha'} X^{v,\alpha,w,\alpha'}_{\vec{q}} \gamma^{\alpha\dagger}_{v,\vec{q}} \gamma^{\alpha'\dagger}_{w,-\vec{q}} + h.c. \\
& + \frac{1}{A} \sum_{\vec{q},\vec{p}}^{v,\alpha,w,\alpha',t,\beta} Y^{v,\alpha,w,\alpha',t,\beta}_{\vec{q},\vec{p}} \gamma^{\alpha\dagger}_{v,\vec{q}} \gamma^{\alpha'\dagger}_{w,-\vec{q}} \gamma^{\beta\dagger}_{t,\vec{p}} \gamma^{\beta}_{t,\vec{p}} + h.c., \\
M_z = & -g_s \mu_b \sum_i b^{\dagger}_{i\alpha} \frac{\sigma^z_{\alpha\beta}}{2} b_{i\beta} \\
& \to \sum_{\vec{q}}^{v,\alpha,w,\alpha'} Z^{v,\alpha,w,\alpha'}_{\vec{q}} \gamma^{\alpha\dagger}_{v,\vec{q}} \gamma^{\alpha'\dagger}_{w,-\vec{q}} + h.c..
\end{aligned} \tag{4.26}$$

The objects $X^{v,\alpha,w,\alpha'}_{\vec{q}}$, $Y^{v,\alpha,w,\alpha',t,\beta}_{\vec{q},\vec{p}}$, $Z^{v,\alpha,w,\alpha'}_{\vec{q}}$ are determined by the Bogoliubov transformation from Eq.(4.9) to Eq.(4.22).

Plugging in Eq.(4.1), one finds

$$\delta_{\hat{n}} \alpha(\omega) = \frac{8\pi\mu_r}{\epsilon_0 c^2 d} \mathrm{Re}[I(\omega)], \tag{4.27}$$

where

$$I(\omega) = \frac{\omega}{A} \sum_{\vec{q}}^{v,\alpha,w,\alpha'} \left[X_{\vec{q}}^{v,\alpha,w,\alpha'} + \frac{1}{A} \sum_{\vec{p}}^{t,\beta} Y_{\vec{q},\vec{p}}^{v,\alpha,w,\alpha',t,\beta} \cdot g_{t,\vec{p}} \right]^*$$

$$\cdot Z_{\vec{q}}^{v,\alpha,w,\alpha'} (1 + g_{v,\vec{q}} + g_{w,-\vec{q}}) \cdot \delta(E_{v,\vec{q}} + E_{w,-\vec{q}} - \hbar\omega). \tag{4.28}$$

Here the bosonic factor $(1 + g_{v,\vec{q}} + g_{w,-\vec{q}})$ is well anticipated from the golden rule. The factor $g_{t,\vec{p}}$ appears because of the quartic interactions in $\vec{P}$ in Eq.(4.8).

It is a good moment to study the symmetry property of $I(\omega)$. In thermal equilibrium, it is straightforward to see that the inversion symmetry alone dictates $I(\omega) = 0$, while time-reversal symmetry alone allows a nonzero imaginary part of $I(\omega)$ (giving rise to the well-known natural circular dichroism in noncentrosymmetric systems).

Next we consider the effect of nonequilibrium occupation $\delta g_{u,\vec{k}}$ in Eq.(4.24). Expanding Eq.(4.28) gives three contributions, $I = I^{(A)} + I^{(B)} + I^{(C)}$:

$$\begin{aligned} I^{(A)}(\omega) &\propto X^* \cdot Z \cdot (\delta g_{v,\vec{q}} + \delta g_{w,-\vec{q}}), \\ I^{(B)}(\omega) &\propto Y^* \cdot Z \cdot g^0_{t,\vec{p}} \cdot (\delta g_{v,\vec{q}} + \delta g_{w,-\vec{q}}), \\ I^{(C)}(\omega) &\propto Y^* \cdot Z \cdot \delta g_{t,\vec{p}} \cdot (1 + g^0_{v,\vec{q}} + g^0_{w,-\vec{q}}). \end{aligned} \tag{4.29}$$

While the inversion symmetry allows all these contributions, the time-reversal symmetry only allows their real parts: the directional dichroism. In addition, in the special situation that $v = w$, namely if the created two spinons are in the same band, obviously $I^{(A)}(\omega) = I^{(B)}(\omega) = 0$ due to Eq.(4.25) and only $I^{(C)}(\omega)$ is nonzero.

Focusing on the low temperature/energy TNDD spectroscopy, one may consider the contribution $v = w = t = 1$ from the lowest energy band only (see Fig.4.2 for a plot of the

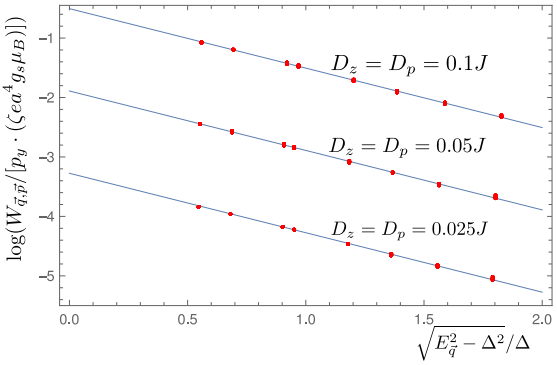


Figure 4.4: The fit $\log(W_{\vec{q},\vec{p}}/[p_y \cdot (\zeta e a^4 g_s \mu_B)]) = \log(u_0) - \sqrt{E^2_{1,\vec{q}} - \Delta^2}/\Delta$ (i.e., Eq.(4.31) with $\vec{u} = u\hat{y} = u_0 \zeta e a^4 g_s \mu_B \hat{y}$) with only one fitting parameter u_0. In each case 696 data points with both $\sqrt{E^2_{1,\vec{p}} - \Delta^2}/\Delta$ and $\sqrt{E^2_{1,\vec{q}} - \Delta^2}/\Delta$ between 0.5 and 1.7 are plotted. Since many data points are related by the lattice symmetry and/or share the same momentum $\vec{q}$ (but different $\vec{p}$), the visibly different data points are much fewer. We set $A = 1$, and consider three cases of different SOC strength: case-(a): $D_z = D_p = 0.025J$ (and $\mu = -1.752J$); case-(b): $D_z = D_p = 0.05J$ (and $\mu = -1.765J$); case-(c) $D_z = D_p = 0.1J$ (and $\mu = -1.792J$). Notice that for each case the chemical potential μ is tuned so that the spinon gap is fixed to be $\Delta = 0.16J$. As shown in this figure, we numerically find that $u_0 = 0.0378$ in case-(a), $u_0 = 0.151 = 0.0378 \cdot 3.99$ in case-(b), and $u_0 = 0.603 = 0.151 \cdot 3.99$ in case-(c). The scaling $u_0 \propto (D/J)^2$ is confirmed.

band structure), and compute $I(\omega) = I^{(C)}(\omega)$ analytically. In this case:

$$I^{(C)}(\omega) = \frac{\omega}{A^2} \sum_{\vec{q},\vec{p}} W_{\vec{q},\vec{p}} \cdot \delta g_{1,\vec{p}} (1 + 2g^0_{1,\vec{q}}) \delta(2E_{1,\vec{q}} - \hbar\omega),$$

$$\text{where } W_{\vec{q},\vec{p}} \equiv \sum_{\alpha,\alpha',\beta} (Y^{1,\alpha,1,\alpha',1,\beta}_{\vec{q},\vec{p}})^* \cdot Z^{1,\alpha,1,\alpha'}_{\vec{q}}. \tag{4.30}$$

$W_{\vec{q},\vec{p}}$ is a real function satisfying $W_{\vec{q},\vec{p}} = -W_{-\vec{q},-\vec{p}}$ due to the inversion symmetry. Taylor expanding near the Γ-point, to the leading order one expects: $W_{\vec{q},\vec{p}} \approx \vec{u} \cdot \vec{p} + \vec{v} \cdot \vec{q}$. In fact, interestingly, we numerically found that $W_{\vec{q},\vec{p}}$ can be well described as

$$W_{\vec{q},\vec{p}} = (\vec{u} \cdot \vec{p})\ e^{-\sqrt{E^2_{1,\vec{q}} - \Delta^2}/\Delta} \tag{4.31}$$

in the momentum regime where the relativistic dispersion Eq.(4.10) holds (see Fig.4.4 for details). We do not attempt to analytically justify Eq.(4.31) here since it deviates from the main purpose of this chapter. Eq.(4.30,4.31) then lead to:

$$I^{(C)}(\omega) = \frac{\omega}{A^2} \sum_{\vec{p}} (\vec{u} \cdot \vec{p}\, \delta g_{1,\vec{p}})$$
$$\cdot \sum_{\vec{q}} e^{-\sqrt{E^2_{1,\vec{q}} - \Delta^2}/\Delta} (1 + 2g^0_{1,\vec{q}}) \delta(2E_{1,\vec{q}} - \hbar\omega). \tag{4.32}$$

Crystal symmetry and dimensional analysis show that $\vec{u} = u\hat{y} = u_0 \zeta e a^4 g_s \mu_B \hat{y}$, consistent with the η_2 response in Eq.(4.5). The dimensionless number u_0 is expect to be $\sim (D/J)^2$ and can be determined numerically (see Fig.4.4 for details).

With Eq.(4.24,4.27,4.10,4.32) the low temperature/energy TNDD response can be computed within our mean-field treatment:

$$\delta_{\hat{y}}\alpha(\omega) = \mathcal{C} \cdot \left[1 + 2g^0(\hbar\omega/2)\right] \cdot (k_B T)^3$$
$$\cdot \left[3G_3(z) - 3\ln z \cdot G_2(z) + (\ln z)^2 G_1(z)\right]$$
$$\cdot e^{-\sqrt{(\hbar\omega/2)^2 - \Delta^2}/\Delta} \cdot \hbar\omega \cdot JDOS(\hbar\omega) \cdot \frac{\nabla_y T \cdot \tau \cdot v}{T}. \tag{4.33}$$

This is just the Eq.(4.11) in the main text.

We can apply the estimate in the previous section to the present example as follows. We firstly estimate $\alpha_{\hat{n}}$ due to the electric dipole processes following the golden rule:

$$\begin{aligned}\alpha_{\hat{n}}(\omega) &\sim \frac{2}{n_r\epsilon_0 c}\frac{2\pi}{\hbar}(\zeta e a)^2[1+2g^0(\hbar\omega/2)]\hbar\omega\cdot JDOS(\hbar\omega)\\ &= \frac{16\pi^2}{n_r}\boldsymbol{\alpha}\zeta^2 a^2[1+2g^0(\hbar\omega/2)]\hbar\omega\cdot JDOS(\hbar\omega),\end{aligned} \tag{4.34}$$

where n_r is material's relative refractive index. For the situation with $k_BT\sim J\sim\frac{\hbar v}{a}$ and $\hbar\omega\sim 2\Delta$, Eq.(4.11,4.34) give the dimensionless ratio $TNDD(\omega)$ in Eq.(4.2):

$$TNDD(\omega)\sim\frac{\boldsymbol{\alpha}a_0}{\zeta a}\cdot\boldsymbol{\mathcal{u}}_0\frac{\nabla_y T\cdot\tau\cdot v}{T}, \tag{4.35}$$

confirming the estimate Eq.(4.3) since $\boldsymbol{\mathcal{u}}_0\propto(D/J)^2$.

Finally, we would like to remark on the validity of the mean-field treatment. Although we performed the calculation within the mean-field approach, the main component of the calculation (Eq.(4.29,4.30) in App.4.4.3) is justified as long as the quasiparticle description is valid. These microscopic contributions to TNDD can be written down phenomenologically as a low quasiparticle-density expansion, up to the second order $\propto g_{\vec{p}}\cdot g_{\vec{q}}$. Some other components of the calculation (e.g., the matrix element behavior Eq.(4.31)) may receive corrections moving beyond the mean-field approximation, but these would not change the result of TNDD response qualitatively.

Bibliography

[1] P. Anderson, Materials Research Bulletin **8**, 153 (1973).

[2] P. A. Lee, Journal of Physics: Conference Series **529**, 012001 (2014).

[3] Y. Zhou, K. Kanoda, and T.-K. Ng, Rev. Mod. Phys. **89**, 025003 (2017).

[4] L. Savary and L. Balents, Reports on Progress in Physics **80**, 016502 (2016).

[5] C. Broholm, R. J. Cava, S. A. Kivelson, D. G. Nocera, M. R. Norman, and T. Senthil, Science **367** (2020), 10.1126/science.aay0668,

[6] X.-G. Wen, Phys. Rev. B **65**, 165113 (2002).

[7] X.-G. Wen, Rev. Mod. Phys. **89**, 041004 (2017).

[8] M. P. Shores, E. A. Nytko, B. M. Bartlett, and D. G. Nocera, Journal of the American Chemical Society **127**, 13462 (2005),

[9] M. R. Norman, Rev. Mod. Phys. **88**, 041002 (2016).

[10] Y. Kasahara, T. Ohnishi, Y. Mizukami, O. Tanaka, S. Ma, K. Sugii, N. Kurita, H. Tanaka, J. Nasu, Y. Motome, T. Shibauchi, and Y. Matsuda, Nature **559**, 227 (2018).

[11] M. Hermele, M. P. A. Fisher, and L. Balents, Phys. Rev. B **69**, 064404 (2004).

[12] M. J. P. Gingras and P. A. McClarty, Reports on Progress in Physics **77**, 056501 (2014).

[13] Or time-reversal symmetry combined with a spatial translation such as in an antiferromagnet

[14] R. Fuchs, The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics 11, 647 (1965), https://doi.org/10.1080/14786436508224252.

[15] D. Szaller, S. Bord´acs, V. Kocsis, T. R˜o om, U. Nagel, and I. K´ezsm´arki, Phys. Rev. B **89**, 184419 (2014).

[16] In general NDD receives contributions from higher order multipole processes. [66] However in the context of Mott insulators the electric-dipole-magnetic-dipole contribution Eq.4.1 dominates.

[17] J. Goulon, A. Rogalev, C. Goulon-Ginet, G. Benayoun, L. Paolasini, C. Brouder, C. Malgrange, and P. A. Metcalf, Phys. Rev. Lett. **85**, 4385 (2000).

[18] M. Kubota, T. Arima, Y. Kaneko, J. P. He, X. Z. Yu, and Y. Tokura, Phys. Rev. Lett. **92**, 137401 (2004).

[19] T. Arima, Journal of Physics: Condensed Matter 20, 434211 (2008).

[20] I. K´ezsm´arki, N. Kida, H. Murakawa, S. Bord´acs, Y. Onose, and Y. Tokura, Phys. Rev. Lett. **106**, 057403 (2011).

[21] Y. Takahashi, R. Shimano, Y. Kaneko, H. Murakawa, and Y. Tokura, Nature Physics **8**, 121 (2012).

[22] Y. Okamura, F. Kagawa, M. Mochizuki, M. Kubota, S. Seki, S. Ishiwata, M. Kawasaki, Y. Onose, and Y. Tokura, Nature Communications **4**, 2391 (2013).

[23] I. Kzsmrki, D. Szaller, S. Bordcs, V. Kocsis, Y. Tokunaga, Y. Taguchi, H. Murakawa, Y. Tokura, H. Engelkamp, T. Rm, and U. Nagel, Nature Communications **5**, 3203 (2014).

[24] S. Toyoda, N. Abe, S. Kimura, Y. H. Matsuda, T. Nomura, A. Ikeda, S. Takeyama, and T. Arima, Phys. Rev. Lett. **115**, 267207 (2015).

[25] Y. Tokura and N. Nagaosa, Nature Communications **9**, 3740 (2018).

[26] P. Jung and A. Rosch, Phys. Rev. B **75**, 245104 (2007).

[27] I. Affleck and J. B. Marston, Phys. Rev. B **37**, 3774 (1988).

[28] M. Hermele, T. Senthil, M. P. A. Fisher, P. A. Lee, N. Nagaosa, and X.-G. Wen, Phys. Rev. B **70**, 214437 (2004).

[29] Y. Ran, M. Hermele, P. A. Lee, and X.-G. Wen, Phys. Rev. Lett. **98**, 117205 (2007).

[30] O. I. Motrunich, Phys. Rev. B **72**, 045105 (2005).

[31] S.-S. Lee and P. A. Lee, Phys. Rev. Lett. **95**, 036403 (2005).

[32] Similar to a thermal transport experiment, if the temperature of the system is far below the magnetic excitation energy, a temperature gradient would not efficiently affect the magnetic excitation distributions and would not lead to a sizable TNDD.

[33] L. N. Bulaevskii, C. D. Batista, M. V. Mostovoy, and D. I. Khomskii, Phys. Rev. B **78**, 024402 (2008).

[34] We only consider the contribution from the spin magnetic moment in this chapter. The orbital magnetic moment in a Mott insulator is a spin-singlet but is much smaller than the spin magnetic moment, by a factor of $(t/U)^2$ in the (t/U)-expansion. [33, 67]

[35] Y. Tokiwa, T. Yamashita, D. Terazawa, K. Kimura, Y. Kasahara, T. Onishi, Y. Kato, M. Halim, P. Gegenwart, T. Shibauchi, S. Nakatsuji, E.-G. Moon, and Y. Matsuda, Journal of the Physical Society of Japan **87**, 064702 (2018).

[36] G. Jackeli and G. Khaliullin, Phys. Rev. Lett. **102**, 017205 (2009).

[37] J. c. v. Chaloupka, G. Jackeli, and G. Khaliullin, Phys. Rev. Lett. **105**, 027204 (2010).

[38] H. Takagi, T. Takayama, G. Jackeli, G. Khaliullin, and S. E. Nagler, Nature Reviews Physics **1**, 264 (2019).

[39] Y. Singh and P. Gegenwart, Phys. Rev. B **82**, 064412 (2010).

[40] J. A. Sears, M. Songvilay, K. W. Plumb, J. P. Clancy, Y. Qiu, Y. Zhao, D. Parshall, and Y.-J. Kim, Phys. Rev. B **91**, 144420 (2015).

[41] R. D. Johnson, S. C. Williams, A. A. Haghighirad, J. Singleton, V. Zapf, P. Manuel, I. I. Mazin, Y. Li, H. O. Jeschke, R. Valent´ı, and R. Coldea, Phys. Rev. B **92**, 235119 (2015).

[42] H. B. Cao, A. Banerjee, J.-Q. Yan, C. A. Bridges, M. D. Lumsden, D. G. Mandrus, D. A. Tennant, B. C. Chakoumakos, and S. E. Nagler, Phys. Rev. B **93**, 134423 (2016).

[43] M. Elhajal, B. Canals, and C. Lacroix, Phys. Rev. B **66**, 014422 (2002).

[44] A. C. Potter, T. Senthil, and P. A. Lee, Phys. Rev. B **87**, 245106 (2013).

[45] Generally the polarization operator contains spin-triplet terms similar to DM interactions. Here for simplicity we only consider spin-singlet terms which dominate in the weak spin-orbit coupling limit.

[46] A. H. MacDonald, S. M. Girvin, and D. Yoshioka, Phys. Rev. B **37**, 9753 (1988).

[47] ζ also receives contribution from the magneto-elastic coupling. For a typical transition metal Mott insulator, this contribution to polarization is similar in size as the contribution from the t/U-expansion [33, 44].

[48] S. Yan, D. A. Huse, and S. R. White, Science **332**, 1173 (2011).

[49] S. Depenbrock, I. P. McCulloch, and U. Schollw¨ock, Phys. Rev. Lett. **109**, 067201 (2012).

[50] Y. Iqbal, F. Becca, S. Sorella, and D. Poilblanc, Phys. Rev. B **87**, 060405 (2013).

[51] H. J. Liao, Z. Y. Xie, J. Chen, Z. Y. Liu, H. D. Xie, R. Z. Huang, B. Normand, and T. Xiang, Phys. Rev. Lett. **118**, 137202 (2017).

[52] Y.-C. He, M. P. Zaletel, M. Oshikawa, and F. Pollmann, Phys. Rev. X **7**, 031020 (2017).

[53] S. Sachdev, Phys. Rev. B **45**, 12377 (1992).

[54] D. P. Arovas and A. Auerbach, Phys. Rev. B **38**, 316 (1988).

[55] N. Read and S. Sachdev, Phys. Rev. Lett. **66**, 1773 (1991).

[56] S. Sachdev and N. Read, International Journal of Modern Physics B **05**, 219 (1991).

[57] Notice that a single spinon excitation is not gauge invariant and does not contribute to physical responses

[58] Y.-M. Lu, Y. Ran, and P. A. Lee, Phys. Rev. B **83**, 224413 (2011).

[59] Y.-M. Lu, G. Y. Cho, and A. Vishwanath, Phys. Rev. B **96**, 205150 (2017).

[60] R. Nandkishore and D. A. Huse, Annual Review of Condensed Matter Physics **6**, 15 (2015).

[61] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, Rev. Mod. Phys. **91**, 021001 (2019).

[62] M. Levin and T. Senthil, Phys. Rev. B **70**, 220403 (2004).

[63] D. V. Pilon, C. H. Lui, T. H. Han, D. Shrekenhamer, A. J. Frenzel, W. J. Padilla, Y. S. Lee, and N. Gedik, Phys. Rev. Lett. **111**, 127401 (2013).

[64] A. Little, L. Wu, P. Lampen-Kelley, A. Banerjee, S. Patankar, D. Rees, C. A. Bridges, J.-Q. Yan, D. Mandrus, S. E. Nagler, and J. Orenstein, Phys. Rev. Lett. **119**, 227201 (2017).

[65] F. Wang and A. Vishwanath, Phys. Rev. B **74**, 174423 (2006).

[66] Y. Gao and D. Xiao, Phys. Rev. Lett. **122**, 227402 (2019).

[67] O. I. Motrunich, Phys. Rev. B **73**, 155115 (2006).